Modeling of spatial extremes in environmental data science: Time to move away from max-stable processes

Raphaël Huser[1], Thomas Opitz[2], and Jennifer Wadsworth[3]

January 30, 2024

Abstract. Environmental data science for spatial extremes has traditionally relied heavily on max-stable processes. Even though the popularity of these models has perhaps peaked with statisticians, they are still perceived and considered as the 'state-of-the-art' in many applied fields. However, while the asymptotic theory supporting the use of max-stable processes is mathematically rigorous and comprehensive, we think that it has also been overused, if not misused, in environmental applications, to the detriment of more purposeful and meticulously validated models. In this paper, we review the main limitations of max-stable process models, and strongly argue against their systematic use in environmental studies. Alternative solutions based on more flexible frameworks using the exceedances of variables above appropriately chosen high thresholds are discussed, and an outlook on future research is given, highlighting recommendations moving forward and the opportunities offered by hybridizing machine learning with extreme-value statistics.

Impact statement. This position paper reviews the severe limitations of max-stable processes for environmental extreme data science, and discusses more appropriate alternative statistical frameworks for the modeling of spatial extremes that have emerged recently. Use of machine learning and AI methods in spatial extreme-value modeling and inference is also discussed, and seven key recommendations to push the field forward are given.

[1]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

[2]INRAE, Biostatistics and Spatial Processes (BioSP, UR546), 228 Route de l'Aérodrome, 84914 Avignon, France. E-mail: thomas.opitz@inrae.fr

[3]Department of Mathematics and Statistics, Fylde College, Lancaster University, Lancaster, United Kingdom. E-mail: j.wadsworth@lancaster.ac.uk

1 Introduction

With the rise of statistical machine learning that marks the 'data science revolution' (Donoho, 2017), and the increasing availability of massive high-quality environmental data products based on observation and simulation (e.g., large climate model ensembles and reanalysis data, remote sensing, wide in-situ observation networks, mobile sensors or citizen-science data), the relevance of traditional statistical models is at stake more than ever. This is especially true with the modeling and prediction of environmental extreme events, where assumptions are crucial for accurate risk assessment and mitigation, and where applied findings are of key societal importance on a global scale (IPCC, 2023). On the one hand, sophisticated models that are well supported by probability theory are desired, in order to provide sound extrapolation into the tail of the distribution. On the other hand, practical considerations should guide the model construction to ensure that it can efficiently utilize available data and provide the answers we need to appropriately address the specific scientific problem at hand. In particular, the variability of processes along the space and time dimensions usually plays a key role in environmental science, and a given statistical model should capture the most important marginal and dependence features of the data, such as spatio-temporal trends and non-stationarity, non-Gaussianity, and subasymptotic forms of tail dependence. Such models should also enable fast-enough inference, which includes model fitting, validation, simulation, and prediction. The speed at which a statistical extreme-value analysis must be performed and the amount of human and material resources required to achieve it depend strongly on the context; while spending months or years could be acceptable for academic purposes or for retrospective studies, it is crucial in some cases to do it within just a few days or weeks (as, e.g., with rapid extreme-event attribution studies to respond to the media about the role of anthropogenic forcings in the occurrence of a recent catastrophic event; see Stott et al., 2004; Risser and Wehner, 2017) or even 'online' (as, e.g., with operational early-warning systems predicting natural hazards in real time, where the safety of people or infrastructure is at risk; see Nguyen et al., 2023). Oftentimes, however, these different requirements are at odds with each other: popular spatial model classes arising from asymptotic extreme-value theory are often computationally prohibitive due to their intricate probabilistic structure, or subject to important practical restrictions that hamper their widespread application in real operational settings, where data are often big and complex. In this paper, we recall that an asymptotic motivation should never supersede applied scientific considerations and proper model checking. Importantly, we argue that the class of <u>max-stable</u> <u>processes</u>, characterized by extreme-value copulas, whose practical usage by statisticians and cli-

mate scientists has been multiplied since the article of Padoan et al. (2010), not only has many severe built-in limitations, but also fails to address the basic purpose it was made for: namely, to provide a suitable statistical framework for modeling spatial extremes and estimating small joint tail probabilities (or, similarly, high return levels of spatial aggregates) far beyond observed levels. The enthusiasm about, and adoption of, max-stable process models in environmental studies—for which we are partly responsible—is due to their solid theoretical foundation, which makes them appear as 'natural' models to use, the fact that the extreme-value community has traditionally been more theory-oriented while being somewhat 'detached' from concrete issues arising in real applications, and the availability of convenient user-friendly software such as the R package SpatialExtremes (Ribatet, 2022) to fit and simulate these models. While (part of) the statistics of extremes community has already realized some of their limitations (Davison et al., 2019; Huser and Wadsworth, 2022) and started to develop alternative modeling strategies that transcend the classical framework (Wadsworth and Tawn, 2012; Opitz, 2016; de Fondeville and Davison, 2018; Huser and Wadsworth, 2019; Engelke and Hitz, 2020; Huser et al., 2021; Wadsworth and Tawn, 2022; Castro-Camilo et al., 2022), max-stable processes and extreme-value copulas still continue to be used in many spatial data applications and considered in simulation studies as the 'default' option.

This paper aims to openly discuss the known deficiencies of max-stable processes, and strongly encourage statisticians and climate scientists to move away from them in real applications unless better alternative solutions are not available. We argue that, as a community, it is now time to reflect and act upon the lessons learned over the past two decades, and move on with the broader adoption of more recent, flexible, efficient, and pragmatic solutions for the modeling of extremes in operational risk assessment studies and, more generally, environmental data science.

The rest of the paper is organized as follows. In Section 2, we review max-stable processes and their main limitations. In Section 3, we discuss alternative modeling strategies. In Section 4, we conclude with some final remarks, an outlook on the future of environmental extreme data science, with a particular view on advances at the interface between statistics of extremes and modern machine learning, and we also list several recommendations moving forward.

2 Max-stable processes: a restrictive tool for the wrong problem?

The theoretical foundations underpinning max-stable processes start with the wish to generalize univariate extreme-value theory (Davison and Huser, 2015) to the spatial context. Consider a sequence of independent and identically distributed random processes, $Y_1(\boldsymbol{s}), Y_2(\boldsymbol{s}), \ldots$, defined over

a spatial domain $\mathcal{S}$. Extreme-value theory states that under broad conditions, the only possible limits, $Z(\boldsymbol{s})$, for the process of pointwise maxima, $Z_n(\boldsymbol{s}) = \max\{Y_1(\boldsymbol{s}), \ldots, Y_n(\boldsymbol{s})\}$ as $n \to \infty$, when appropriately affinely renormalized, are max-stable processes. This result implies that all univariate margins of $Z(\boldsymbol{s})$ follow the generalized extreme-value (GEV) distribution, while all finite-dimensional margins are characterized by an extreme-value copula (Davison et al., 2012; Segers, 2012). This asymptotic characterization (as the block size n tends to infinity) has been the principal argument for fitting max-stable models and extreme-value copulas in practice (with fixed and finite n). While the max-stable process theory was established in the 1980s, their popularity started to grow with Schlather (2002) who showed how to construct max-stable models with realistic-looking realizations based on de Haan (1984)'s spectral representation, and their use in environmental applications was later boosted by Padoan et al. (2010) who proposed a method of inference for spatial max-stable models based on pairwise likelihoods. Davison et al. (2012) further advocated their use against other natural alternatives available at the time. The spectral representation of max-stable processes (de Haan, 1984; Schlather, 2002) essentially states that, on the unit Fréchet scale, they can be constructed as

$$Z(\boldsymbol{s}) = \sup_{i \geq 1} \xi_i W_i(\boldsymbol{s}), \tag{1}$$

for a Poisson point process $\{\xi_i\}_{i \geq 1}$ on $(0, \infty)$ with intensity $\xi^{-2}\mathrm{d}\xi$, and independent copies $W_i(\boldsymbol{s})$ of a spatial process $W(\boldsymbol{s})$ satisfying $\mathrm{E}[\max\{W(\boldsymbol{s}), 0\}] = 1$. The unit Fréchet scale is a common standard since it permits simple expression of the finite-dimensional distribution functions as $G(\boldsymbol{z}) = \exp\{-V(\boldsymbol{z})\}$, with V the so-called exponent function, which is homogeneous of order -1, meaning that $t \times V(t\boldsymbol{z}) = V(\boldsymbol{z})$ for any positive t and $\boldsymbol{z}$. This construction principle has led to various max-stable process models, the most popular ones being the Smith (1990) model, the Schlather (2002) model, the extremal-t model (Opitz, 2013), the Brown and Resnick (1977) model (Kabluchko et al., 2009), and the Reich and Shaby (2012) model, which have been widely used in numerous applications. However, max-stable processes have many intrinsic limitations and practical restrictions, and we now summarize the most critical ones. These restrictions are related to the actual definition of max-stable processes, the form of their dependence structure, and the cost of likelihood inference and simulation algorithms.

Definition of max-stable processes A first drawback of max-stable processes goes back to their basic theoretical motivation, and the implicit definition of a spatial extreme event in this framework. By construction, max-stable processes $Z(\boldsymbol{s})$ approximate the distribution of pointwise
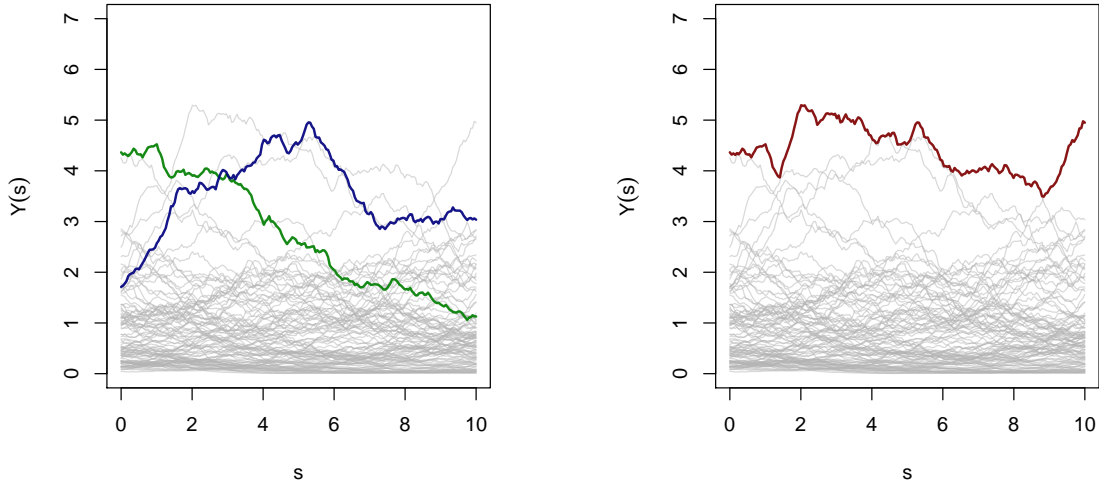
Figure 1: <u>Left:</u> Illustration of underlying random fields $Y_i(\boldsymbol{s})$ in gray, with two of the fields that contribute to the pointwise maximum highlighted in green and blue. <u>Right:</u> Illustration of underlying random fields $Y_i(\boldsymbol{s})$ in gray, with the pointwise maximum $Z_n(\boldsymbol{s})$ highlighted in red. In this case, there is no $j = 1, \ldots, n$ such that $Z_n(\boldsymbol{s}) = Y_j(\boldsymbol{s})$ for all $\boldsymbol{s}$.

maxima $Z_n(\boldsymbol{s})$ for large n. However, while the $Y_i(\boldsymbol{s})$'s represent the original individual (e.g., daily) spatial events, which are directly observable and may or may not be extreme, the block (e.g., yearly) maximum process $Z_n(\boldsymbol{s})$ does not correspond to a real event, unless a single individual process, say $Y_j(\boldsymbol{s})$, is more extreme than <u>all</u> other processes $Y_i(\boldsymbol{s})$, $i \in \{1, \ldots, n\} \setminus \{j\}$, simultaneously at <u>all</u> sites $\boldsymbol{s}$ within the domain $\mathcal{S}$, such that $Z_n(\boldsymbol{s}) = Y_j(\boldsymbol{s})$ for all $\boldsymbol{s}$. This situation rarely occurs in practice, especially for large domains. Therefore, in fitting a max-stable process to realizations of $Z_n(\boldsymbol{s})$, we effectively model artificially-created spatial extreme data that may have little to do with real observations. This is illustrated in Figure 1. An argument used to persist and still fit max-stable processes to extremes of individual observations (e.g., Huser and Davison, 2014) is that they provide, in some sense, an approximation to the joint tail of Y_i itself. Specifically, if a vector $\boldsymbol{Y} \sim F$ has a joint distribution F with unit Fréchet margins attracted to the max-stable distribution $G(\boldsymbol{z}) = \exp\{-V(\boldsymbol{z})\}$, then we have $F(n\boldsymbol{y}) = [\{F(n\boldsymbol{y})\}^n]^{1/n} \approx \{G(\boldsymbol{y})\}^{1/n} = G(n\boldsymbol{y})$, where the last equality uses the homogeneity of V, and the approximation is justified when n is large. However, as is clear from this simple derivation, the max-stable distribution G only provides an accurate approximation to F when <u>all</u> observations are large at the same time, and it does not provide a suitable approximation in the much more common situation where only a subset of variables are extreme. This issue can be partly dealt with by censoring small observations when performing

inference; nevertheless, we argue that, by design, max-stable processes address the 'wrong problem', and that the very statistical framework motivating their practical use is flawed.

The rigidity of stability   Max-stable processes possess a dependence structure that is too rigid for most environmental applications. A max-stable distribution G is one for which G^t (with $G^t(x)$ defined as $\{G(x)\}^t$) remains a valid distribution within the same location-scale family for all $t > 0$. This implies that the dependence structure of max-stable processes is invariant to the maximum operator; in other words, if a specific max-stable process is an appropriate model for annual maxima of a given variable, then the same model also provides an appropriate characterization of the dependence structure for 10-year maxima, 50-year maxima, or even 1000-year maxima. This means that, upon marginal standardization, the dependence patterns of a spatial extreme event do not change with the severity of the event, no matter how extreme it is. This rigidity often contradicts empirical findings, where extreme events tend to become spatially more localized as they become more extreme (Huser and Wadsworth, 2019; Castro-Camilo and Huser, 2020; Zhong et al., 2022). Very few studies which employ max-stable processes actually scrutinize this stability property, but diagnostic tools can be found in Gabda et al. (2012) and Huser et al. (2021), or could be adapted from available multivariate hypothesis testing tools (Bücher and Kojadinovic, 2016).

Models only for limited sub-classes of possible tail structures   The asymptotic world in which max-stable processes live is also 'black and white': max-stable models are indeed always asymptotically dependent unless they are exactly independent, and they lack more nuanced representation of the important case of asymptotic independence. Asymptotic independence means that the limiting dependence structure of the normalized $Z_n(\boldsymbol{s})$ corresponds to the independence copula, yet in practice there is almost always residual positive dependence present in $Z_n(\boldsymbol{s})$ for finite n, even when the limit would be independence. By fitting a max-stable process model to maxima of asymptotically independent data, one incorrectly models this residual dependence as asymptotic dependence, leading to biased extrapolation further into the tail of the distribution. Put differently, a max-stable model fitted to block maxima stemming from data with a weakening tail dependence structure is inevitably misspecified; the effect of this misspecification is that max-stable processes will capture an 'average' strength of dependence and, therefore, tend to slightly underestimate the occurrence probability of joint extreme events at relatively low levels but potentially grossly overestimate them at high levels situated far beyond the observed range of data. Another way to define asymptotic dependence of the process $Y(\boldsymbol{s})$ is to consider the coefficient

$\chi_{\{i,j\}}(u) = \Pr\{Y(\boldsymbol{s}_i) > F_i^{-1}(u), Y(\boldsymbol{s}_j) > F_j^{-1}(u)\}/(1-u)$, where $Y(\boldsymbol{s}_i) \sim F_i$ and $Y(\boldsymbol{s}_j) \sim F_j$. Asymptotic dependence is present if this converges to a positive limit $\chi_{\{i,j\}} > 0$ as $u \to 1$. In other words, the joint tail decay rate is proportional to the marginal tail decay rate, and any dependence at moderate levels never vanishes completely but it remains in the limiting tail. While asymptotic dependence (or independence) is a property that is difficult to verify or test in practice, models allowing for asymptotic independence often have richer tail decay rates than max-stable processes, which are strongly limited due to their focus on the possible asymptotic structures rather than the subasymptotic tail behavior. This difference is crucial in practice, because the flexibility of a model in its joint tail dictates extrapolations to higher levels, and thus impacts risk assessment. Being always asymptotically dependent, max-stable processes will thus have a tendency to overestimate the risk of very large joint extremes. Comparative studies advising against asymptotically dependent models for environmental risk assessment include Bortot et al. (2000) for oceanographic processes and Opitz (2016) and Dawkins and Stephenson (2018) for extreme wind gusts.

Lack of flexible and physically realistic models   Some of the most popular max-stable models are also characterized by unphysical properties. The Schlather and extremal-t max-stable models, for example, are non-ergodic, which implies that they cannot approach full independence between infinitely-distanced sites. With such models, spatial extreme events have a positive probability to be 'infinitely wide' in extent since the $\chi_{\{i,j\}}$ coefficient for $Z(\boldsymbol{s})$ is uniformly bounded below by a positive constant for any two locations $\boldsymbol{s}_i$ and $\boldsymbol{s}_j$ whatever their distance $\|\boldsymbol{s}_i - \boldsymbol{s}_j\|$. On the other hand, the Smith model has overly smooth realizations based on analytical spatial profiles W_i in (1), and the original Reich–Shaby model is a noisy version of the Smith model with artificial non-stationary artefacts. The Brown–Resnick model, which is constructed from intrinsically stationary log-Gaussian processes W_i in (1), seems to be the most physically reasonable one. Nevertheless, the very broad subclass of max-stable processes possessing a positive continuous density in all of their finite-dimensional distributions (including the Brown–Resnick model itself) shares the common drawback that conditional independence implies full independence (Papastathopoulos and Strokorb, 2016). This means that most max-stable models cannot exhibit any interesting Markov structure directly in $Z(\boldsymbol{s})$ but at most in latent variables used to construct the max-stable process (Améndola et al., 2022; Engelke and Hitz, 2020), which is a major impediment to leveraging such Markov structures for efficient inference and validation of max-stable models. This also implies that physically-meaningful stochastic partial differential equation models (see, e.g., Lindgren et al., 2009; Bolin and Wallin, 2020; Zhang et al., 2023b), which commonly lead to conditional independencies

7

and graphical structures when discretized, are directly incompatible with max-stable processes.

Computational complexity Finally, max-stable processes are also notoriously computationally cumbersome to fit and simulate from. Likelihood-based inference is especially challenging, even with advanced supercomputers (Castruccio et al., 2016), because the full likelihood function contains a number of terms that grows super-exponentially fast with the number of observed locations. Therefore, alternative inference solutions are required. Proposed approaches include pairwise likelihoods (Padoan et al., 2010), higher-order composite likelihoods (Genton et al., 2011; Huser and Davison, 2013; Huser et al., 2023), M-estimators (Yuen and Stoev, 2014; Einmahl et al., 2016), a customized stochastic expectation-maximization (EM) algorithm (Huser et al., 2019), distributed inference through a divide-and-conquer strategy (Hector and Reich, 2023), or more recently neural Bayes estimators (Lenzi et al., 2023; Sainsbury-Dale et al., 2023b). However, except for the latter which can be trained offline, these approaches all face a delicate trade-off between computational and statistical efficiency, and they often remain quite expensive to apply in moderate-to-high dimensions in terms of number of spatial locations. Likelihood-based inference can be simplified by including event times in the dataset (Stephenson and Tawn, 2005), but this can lead to bias if the number of locations is large in comparison to the number of replicates over which maxima are taken (Wadsworth, 2015). Nonetheless, the inclusion of event times starts to mimic the paradigm of peaks-over-threshold modeling, which we argue in Section 3 is a more sensible and natural approach.

The stochastic representation (1) involving an infinite number of processes over which the pointwise maximum is taken also makes simulation cumbersome. While various types of approximate and exact simulation algorithms have been developed for certain max-stable families (Schlather, 2002; Oesting et al., 2012; Thibaud and Opitz, 2015; Dombry et al., 2016; Oesting and Strokorb, 2022), conditional simulation remains computationally laborious (Dombry et al., 2013; Oesting and Schlather, 2013) and does not scale well with the dimension. To construct more "generic" stochastic generators for spatial extremes, it is also possible to directly exploit generative artificial-intelligence (AI) techniques from the machine learning literature, such as Generative Adversarial Networks (GANs; see, e.g., Boulaguiem et al., 2022) which bypass the need to specify a parametric dependence structure for extremes at the expense of completely abandoning any known structure, or Variational Autoencoders (VAEs; see, e.g., Lafon et al., 2023; Zhang et al., 2023a), though these recent machine-learning-based approaches are sometimes difficult to train, are always data-hungry, and are often challenging to study from a theoretical perspective, e.g., in terms of the approximation quality of the trained generator, or its ability to accurately reproduce joint tail decay rates.

Discussion   Overall, the difficulties with max-stable processes are 'built-in': they are a direct consequence of their basic definition leading to the complex structure in (1), and the fact that they are not meant to describe extremes of the original individual events. The action of computing block maxima indeed masks information about the timings of events and temporal dependence, and specifically about co-occurrence of maxima at different spatial locations, which has implications for modeling, inference, and simulation. Max-stability arises as an 'asymptotic artefact' resulting from taking the limit of block maxima as the block size n goes to infinity; in practice, however, interest often lies in the original events themselves, rather than maxima. Moreover, even when the modeling of maxima may be desired, the effective block size is often quite moderate in most environmental data due to serial dependence and seasonality, which can create a severe mismatch between theory and practice. Using the limit model for data in this context severely constrains the form of dependence structures that can be obtained, in a way that is unrealistic in most environmental applications, while simultaneously complicating statistical inference significantly.

In the next section, we summarize more recent modeling strategies that bypass several of the above roadblocks by going beyond max-stability, and that are thus better suited for the spatial modeling of extremes.

3   Solutions beyond max-stability

In Section 2, we argued strongly against the continued use of max-stable processes in all but exceptional circumstances. Here we outline the breadth of alternative models for spatially-indexed environmental data, highlighting their relative merits and potential drawbacks.

Peaks-over-threshold vs maxima   Unlike approaches based on block maxima, peaks-over-threshold approaches focus on modeling extremes of the original spatial events that effectively took place. Therefore, they are not only more meaningful from a practical perspective, but they also offer ways to customize the definition of a 'spatial extreme event' to the specific problem of interest. In particular, peaks-over-threshold approaches do not only provide valid probability approximations when all variables are simultaneously large (which rarely occurs in practice), but they can be adapted to events where only a subset of variables are extreme. This is illustrated in Figure 2 for the two main peaks-over-threshold approaches, namely the (generalized) Pareto process (Ferreira and de Haan, 2014; Dombry and Ribatet, 2015) and the spatial conditional extremes process (Wadsworth and Tawn, 2022), which are justified by different asymptotic paradigms.
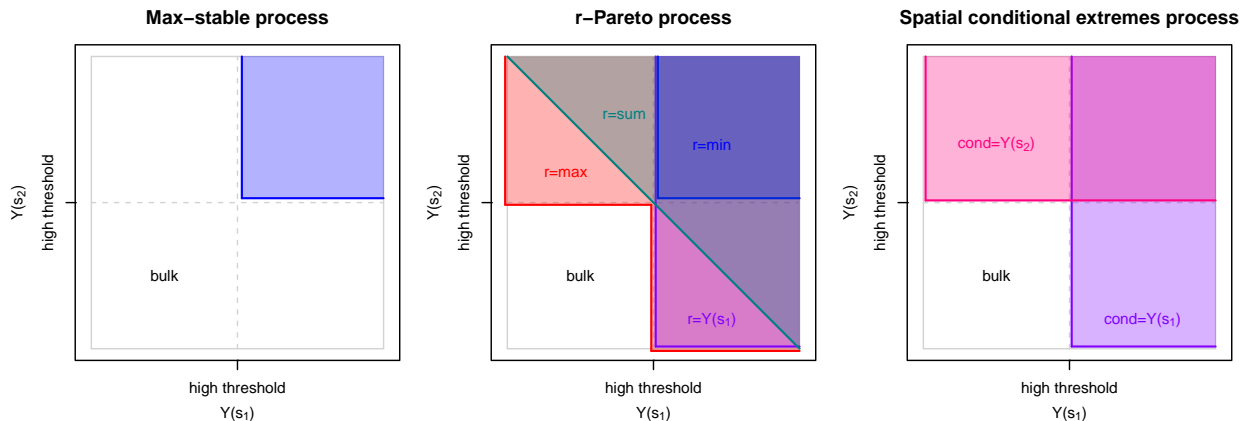
Figure 2: For a pair of variables $\{Y(\boldsymbol{s}_1), Y(\boldsymbol{s}_2)\}$ from various extreme-value models, illustration of the domains (colored areas) over which each model is meant to provide accurate tail probability approximations. <u>Left:</u> max-stable process; <u>Middle:</u> Pareto process for various aggregation functionals, r; <u>Right:</u> Spatial conditional extremes process, for each conditioning variable.

Pareto processes (Ferreira and de Haan, 2014; Dombry and Ribatet, 2015; de Fondeville and Davison, 2018, 2022) are usually viewed as the peaks-over-threshold analogue of max-stable processes. Both classes of models are grounded in the theory of functional regular variation, but Pareto processes are, in principle, applicable to all data which are in some sense extreme. As a result, they possess much simpler likelihoods than max-stable models, as well as making a more efficient use of data. In Figure 1, Pareto processes would represent candidate models for the fields highlighted in green or blue. Slightly different formulations of Pareto processes arise depending on the so-called aggregation (or risk) functional, usually denoted by $r(\cdot)$, used to define a functional extreme event (such as the spatial maximum); see, e.g., de Fondeville and Davison (2018, 2022) for more details.

The spatial conditional extremes model (Wadsworth and Tawn, 2022), on the other hand, is a different peaks-over-threshold approach that only applies to the case where a single variable (at a fixed chosen location) exceeds a high threshold; however, it offers improved tail flexibility and other benefits compared to Pareto processes, as discussed below.

Improved tail flexibility with 'subasymptotic' models   While Pareto processes have simpler likelihoods and permit the use of more data than max-stable processes, they are unfortunately seldom appropriate models in practice when used for tail extrapolation and estimation of small probabilities associated with environmental extreme events that lie far in the upper joint tail. This is because their <u>threshold-stability</u> property, analogous to the max-stability property of max-stable processes, is rarely satisfied by the environmental data available at observed levels of extremity. The
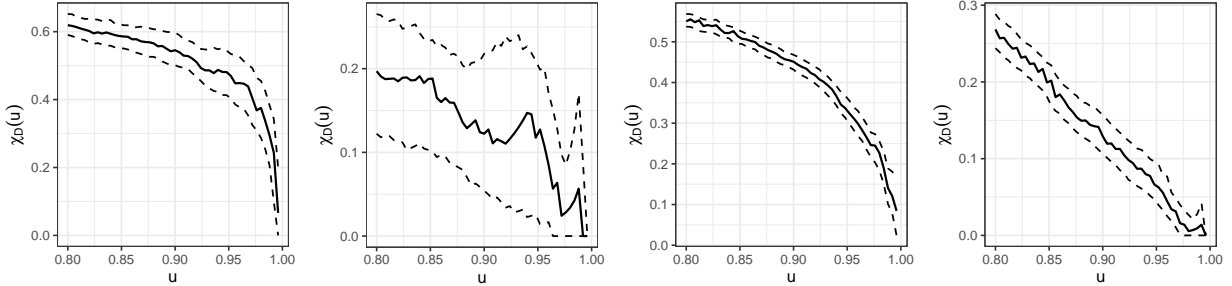
Figure 3: Examples of estimates of $\chi_D(u)$ for four environmental datasets. Solid lines represent point estimates, whereas dashed lines are approximate 95% pointwise confidence intervals based on block bootstrapped estimates. From left to right: (detrended) Red Sea surface temperature data at $d = 144$ locations in the Gulf of Aqaba; gridded conditionally simulated E-OBS Irish summer temperature data at $d = 178$ locations; daily cumulated precipitation at $d = 7$ locations in the Vaucluse 'département' in France; daily mean windspeed at the same $d = 7$ locations in France.

threshold-stability property will never be satisfied by data that exhibit asymptotic independence, and represents a very strong additional assumption over the presence of asymptotic dependence. Let $Y_j = Y(\boldsymbol{s}_j) \sim F_j, j \in D = \{1, \ldots, d\}$, represent a spatial process observed at d locations. A simple way to assess whether this property may hold is to consider the quantity $\chi_D(u)$, defined similarly to $\chi_{\{i,j\}}(u)$ in Section 2, as

$$\chi_D(u) = \Pr\{Y_1 > F_1^{-1}(u), \ldots, Y_d > F_d^{-1}(u)\}/(1 - u), \qquad u \in (0, 1). \tag{2}$$

If a Pareto process is applicable, then for any d there always exist d distinct locations such that $\chi_D(u) \equiv \chi_D > 0$ for all sufficiently large $u < 1$. Our collective experience is that, in contrast, it is almost always the case that $\chi_D(u)$ decreases as $u \to 1$, representing a weakening of spatial dependence at extreme levels. This is illustrated for datasets of sea surface temperature, air temperature, precipitation, and windspeed in Figure 3.

To deal with this deficiency, some authors—including ourselves—have advocated the use of what are often termed subasymptotic models. This terminology is used because such models can be seen to bridge the gap between the world of finite-level data and the 'mythical land' of asymptopia, where models such as max-stable and Pareto processes should be applicable. However, it is perhaps a misnomer, since if data exhibit asymptotic independence, then Pareto processes are not well defined for domains $\mathcal{S}$ composed of an infinite number of locations, and max-stable models offer no benefits. Therefore, aside from conditional extremes models, alternatives such as these are currently the only approach to performing useful extreme value inference.

Broadly, these subasymptotic models are designed to represent flexible forms of tail decay,

11

permitting extrapolation from observed levels to more extreme levels. A typical consideration is that probabilities such as $\Pr\{Y_1 > F_1^{-1}(u), \ldots, Y_d > F_d^{-1}(u)\}$ or $\Pr\{Y_i > F_i^{-1}(u), Y_j > F_j^{-1}(u)\}$ should have flexible forms as $u \to 1$, and these forms fit with the assumptions of regular variation and/or hidden regular variation (Ledford and Tawn, 1996, 1997; Resnick, 2002) to provide some theoretical grounding. Where modeling componentwise maxima remains relevant, we can relax the max-stability property to obtain more realistic dependence structures with constructions such as max-infinitely divisible processes (Huser et al., 2021) or max-mixtures (Wadsworth and Tawn, 2012; Bacro et al., 2016); however, working with such extensions of max-stable processes can further exacerbate the issues of model interpretation and computational cost outlined before.

To date, most subasymptotic models have a random scale construction: extreme data are modeled using the spatial copula of the process $X(\boldsymbol{s}) = RW(\boldsymbol{s})$, where the scalar random variable $R > 0$ is independent of the spatial process W. The relative tail heaviness of R and W, together with the dependence structure of W, provides a rich array of dependence possibilities; see Engelke et al. (2019) for an almost-exhaustive description. Examples of this class of models include Opitz (2016), Huser et al. (2017) and Huser and Wadsworth (2019). With appropriate specification of R and W, Pareto processes also possess a random scale representation (Ferreira and de Haan, 2014). Limitations of simple random scale models include their inability to capture complex dependence structures observed over large domains and a gradual decay of positive dependence to independence at large distances (because of the spatially constant R variable which makes them non-ergodic), as well as the incorporation of non-stationarity. These limitations also make it difficult to adapt such models to the case of spatio-temporal data. Hazra et al. (2022) recently attempted to address these issues by proposing a Gaussian scale mixture model extension, which can capture short-range asymptotic dependence, mid-range asymptotic independence, and long-range exact independence, by replacing R with a suitable spatial process $R(\boldsymbol{s})$; see also Krupskii and Huser (2022).

Conditional spatial extremes model The recently-introduced conditional spatial extremes model (Wadsworth and Tawn, 2022) is another class of models with flexible tail structures, which has been adapted to the spatio-temporal case in Simpson and Wadsworth (2021). These models are based on the assumption of a limiting process for suitably-normalized $Y(\boldsymbol{s})$, conditional upon the event $Y(\boldsymbol{s}_0) > t$, for some conditioning location $\boldsymbol{s}_0$. The formulation permits modelling of both asymptotically dependent and asymptotically independent data, while the most commonly-used version of the likelihood is relatively simple. The 'price' for these two major gains is the necessity of conditioning on a particular location being large, rather than, say, any location being large,

though Wadsworth and Tawn (2022) outline ways in which this can be mitigated if it is an issue.

Considerations for inference As mentioned, the fitting of peaks-over-threshold models via likelihood is much simpler than max-stable models. However, to avoid bias in estimation of the tail properties, it is necessary to fit the models only to extreme data. For generalized Pareto process and subasymptotic models, this is usually done using an appropriately censored likelihood (see, e.g., Huser and Wadsworth, 2022) to avoid influence of small values. Censoring can be done in different ways, either by applying a risk functional such as the maximum to the observation vector and fully censoring or discarding the vectors for which the risk is below a fixed high threshold, or using an approach focusing on marginal exceedances, without the need to define a global risk functional, where any component of the vector that falls below its marginal threshold is censored. Especially in the latter case, this act of censoring makes likelihood evaluation significantly more computationally intensive as it can require calculating numerous potentially high-dimensional integrals of the density function. Likelihood-based inference for such models is therefore typically limited to numbers of observation locations less than about 30, so de Fondeville and Davison (2018) instead advocated using a weighted gradient scoring approach that mimics smooth censoring, while remaining relatively cheap computationally.

The spatial conditional extremes model can usually be fitted to data from hundreds of observation locations since its common variants take the form of a nonstationary and marginally transformed Gaussian process. The different nature of the asymptotics means that an extreme event is one that is extreme at the conditioning location s_0, but that may be large or small elsewhere, which (often) avoids the need for censoring in the likelihood. Simpson et al. (2023) outline extensions to thousands of dimensions via a slight change in formulation and use of Gaussian Markov random fields; see also Vandeskog et al. (2022) for related methodology.

Techniques from the machine learning literature have recently permeated the world of spatial extreme value modeling: Sainsbury-Dale et al. (2023b), Sainsbury-Dale et al. (2023a) and Walchessen et al. (2023) describe the use of neural networks (NNs) for likelihood-free inference on the parameters of spatial models whose likelihoods are costly to evaluate because of high dimensionality. Richards et al. (2023) extend these ideas to incorporate censoring of small values, a key consideration for extremes. These NN-based estimators are 'amortized', meaning that all computational effort is encapsulated in a training period, after which estimates can be obtained in a fraction of a second; this is in comparison to hours or sometimes days for likelihood-based estimates.

Discussion It is almost always the case that interest lies in understanding the extremes of original events, thus it makes sense to model their extremal behavior directly. This usually leads to simpler inference and more flexible classes of models. One reason sometimes cited for preferring a block maximum approach is that the resulting $Z_n(\boldsymbol{s})$ should be independent in time, as the environmental events comprising the block maxima will fall in different years. Although the modeling of original events does lead to more temporal dependence in extremes, it is typically preferable to ignore this when fitting via likelihood and adjust the uncertainty of parameter estimates post-hoc (Fawcett and Walshaw, 2007).

Recent developments in machine-learning-based inference begin to circumvent some computational difficulties for inference with both peaks-over-threshold models and max-stable processes alike. Nonetheless, the conceptual drawbacks of max-stable processes remain, together with potential simulation difficulties. These recent machine-learning-based developments have largely thus far focused on demonstrating how to fit existing models in previously unfeasible scenarios. An exciting possibility is their potential to facilitate generation of flexible new models with desirable tail properties: it is often easier to write down a stochastic representation for a flexible model than to derive (and evaluate) its likelihood function. For example, this could permit specification of and inference on models exhibiting asymptotic dependence at short range, with asymptotic independence and exact independence at longer range. The modeling of spatio-temporal extreme dependence, thus far tackled in relatively few cases, may also be facilitated via this route.

4 Conclusion

The probability theory supporting max-stable processes is rich and should not be despised, as it has contributed significantly to extreme-value theory and led to important advances as well as a better understanding of extremes in stochastic processes; we thus do not question the rigor or historical developments of the theory itself, but rather its relevance in concrete environmental applications. While the systematic use of max-stable processes in environmental studies needs to be gradually phased out, we also do not categorically advocate against the use of asymptotically justified models. Asymptotic theory can indeed be very useful provided the asymptotic paradigm directly responds to a concrete need posed by the applied scientific problem at hand. In particular, peaks-over-threshold approaches should be prioritized over block maxima approaches whenever possible. Pareto processes, and the more recent spatial conditional extremes model, are two possible frameworks that stem from more helpful asymptotic regimes. However, while asymptotic

guarantees are in principle desired for tail extrapolation, we also stress that they should never supersede careful model checking. Oftentimes, more pragmatic solutions (e.g., certain types of random location and scale constructions, or physics-informed models) that display improved sub-asymptotic tail flexibility, physically more realistic properties, or computationally more affordable inference, can be better suited to address the specific scientific problem and should thus not be disregarded. When risk assessment of future extreme events (within a reasonable time frame) is of interest, the incorporation of non-stationary climate change signals (e.g., from climate model outputs under various greenhouse gas emission scenarios) and the development of asymptotic independence models with a flexible joint tail decay rate is more important than focusing on accurately identifying the asymptotic dependence class. Furthermore, when models are intended to be used in operational settings, some accuracy must sometimes be traded for speed of inference; in this context, geostatistical models constructed from Gaussian building blocks and/or based on sparse probabilistic structures, and fast approximate inference and simulation techniques (e.g., based on deep learning), can be particularly helpful.

An interesting direction for future research is to develop models and methods deeply rooted in extreme-value theory that harness the power and computational efficiency of advanced machine learning (e.g., deep extreme quantile regression, deep non-stationary spatial models, neural inference and generative approaches, etc.) to better address problems in environmental data science; machine learning tools can indeed overcome limitations of classical statistical tools designed for estimating few model parameters from datasets of only moderate size. Wikle and Zammit-Mangion (2023) review statistical deep learning methods in classical spatial statistics; with some suitable adjustments, these methods could potentially be adapted to the extreme-value context.

In the future, it would also be interesting to develop unified spatial frameworks for tractably modeling the full data range in a way that offers high flexibility in the lower joint tail, bulk, and upper joint tail, in the same vein as Naveau et al. (2016) in the univariate context. For example, normal mean-variance mixtures allow for both asymptotic dependence and independence as well as for controlling asymmetry in lower- und upper-tail dependence (Zhang et al., 2022). Another related future line of research is to build upon recent advances in the geometric approach for multivariate extremes (Nolde, 2014; Nolde and Wadsworth, 2022). The great benefit of this new asymptotic framework is that it provides a unified representation of multivariate extremes approaches (Wadsworth and Campbell, 2022) and a flexible strategy for the joint modeling of extremes in "all directions"—including the lower and upper tails (Papastathopoulos et al., 2023); see

also Murphy-Barltrop et al. (2024) for related methodology. Extending this approach to the spatial context and to capturing both bulk and tail behaviors simultaneously is an interesting area of investigation. Capturing the full range of values is particularly important for modeling compound extremes (AghaKouchak et al., 2020; Zscheischler et al., 2020) defined as the combination of conditions (over space, time or several variables) leading to extreme impacts but where the contributing events are not necessarily extreme individually. With spatial data increasingly available on regular grids at relatively high resolution, such as climate model output and remote sensing data, the study of geometric properties of the pixellated exceedance sets at increasingly high threshold levels provides another line of research toward better understanding and modeling of the behavior of joint extremes and could offer methods that scale well to very large data volumes (Cotsakis et al., 2023).

Although highly promising, neural likelihood-free parameter inference, mentioned in both Sections 2 and 3, still requires several developments to replace traditional likelihood-based inference. For example, theoretical guarantees on the accuracy of neural estimators in terms of the chosen NN architecture and number of training samples remain to be established. Furthermore, while uncertainty quantification can be handled via the bootstrap, for example, more needs to be understood about its properties. The effects of model misspecification also need to be studied more comprehensively: with likelihood-based inference, it is known that parameter point estimates are robust to certain types of misspecification, and that adjustments can be made for properly handling uncertainty (White, 1982; Chandler and Bate, 2007). Model selection and comparison techniques, which have often relied on log-likelihood values, also require attention. Finally, the estimation of a relatively large numbers of parameters in flexible models for spatial tail dependence remains challenging. While some of these challenges can be addressed by adopting some recent amortized fully-Bayes neural methods, such as BayesFlow (Radev et al., 2020, 2023b) or JANA (Radev et al., 2023a), their use in extreme-value applications remains to be carefully investigated.

Our main recommendations moving forward can be summarized with the acronym MACHINE, which highlights seven 'golden rules' for extreme-value analyses and the special role that machine learning is likely to play in the future of environmental extreme data science:

- **M**ove away from max-stable processes;

- **A**dopt a peaks-over-threshold approach or a unified bulk-tail model whenever possible;

- **C**apture subasymptotic behavior rather than focusing on the asymptotic structure and the dichotomy between asymptotic dependence and independence;

- **H**arness specialized models with a sparse and numerically convenient probabilistic structure for speed and interpretation;

- **I**ncorporate physics/climate knowledge into probability models as much as possible;

- **N**ever-prioritize asymptotic justification over careful model checking;

- **E**mbrace modern machine learning and AI methods to enhance the modeling and inference of extreme events in complex settings.

'Starting the MACHINE' (or keeping it on) will be key, in our opinion, to remain relevant and maximize the impact of extreme-value theory across statistics and applied environmental sciences, especially in the face of the escalating challenges posed by today's world of extreme climate-driven events.

Acknowledgments

We warmly thank the Environmental Data Science (EDS) editorial board and D. Nychka from the advisory board for inviting us to write this opinion piece.

Authors contributions

All authors contributed equally to this work.

Competing interest statement

The authors have no competing interests to declare.

Data availability statement

Data to reproduce Figure 3 can be freely obtained upon reasonable request from the authors or through the following websites: https://data.marine.copernicus.eu/product/SST_GLO_SST_L4_NRT_OBSERVATIONS_010_001/description (Red Sea surface temperature data); https://www.ecad.eu/download/ensembles/download.php (E-OBS Irish temperature data); and https://meteo.data.gouv.fr/form (French precipitation and windspeed data).

Funding statement

References

AghaKouchak, A., Chiang, F., Huning, L. S., Love, C. A., Mallakpour, I., Mazdiyasni, O., Moftakhari, H., Papalexiou, S. M., Ragno, E. and Sadegh, M. (2020) Climate extremes and compound hazards in a warming world. Annual Review of Earth and Planetary Sciences **48**, 519–548.

Améndola, C., Klüppelberg, C., Lauritzen, S. and Tran, N. M. (2022) Conditional independence in max-linear Bayesian networks. The Annals of Applied Probability **32**, 1–45.

Bacro, J.-N., Gaetan, C. and Toulemonde, G. (2016) A flexible dependence model for spatial extremes. Journal of Statistical Planning and Inference **172**, 36–52.

Bolin, D. and Wallin, J. (2020) Multivariate type g matérn stochastic partial differential equation random fields. Journal of the Royal Statistical Society: Series B **82**, 215–239.

Bortot, P., Coles, S. and Tawn, J. A. (2000) The multivariate Gaussian tail model: An application to oceanographic data. Journal of the Royal Statistical Society: Series C (Applied Statistics) **49**, 31–049.

Boulaguiem, Y., Zscheischler, J., Vignotto, E., van der Wiel, K. and Engelke, S. (2022) Modeling and simulating spatial extremes by combining extreme value theory with generative adversarial networks. Environmental Data Science **1**, e5.

Brown, B. M. and Resnick, S. I. (1977) Extreme values of independent stochastic processes. Journal of Applied Probability **14**, 732–739.

Bücher, A. and Kojadinovic, I. (2016) Extreme value modeling and risk analysis: Methods and applications, chapter An overview of nonparametric tests of extreme-value dependence and of some related statistical procedures, pp. 377–398. CRC Press.

Castro-Camilo, D. and Huser, R. (2020) Local likelihood estimation of complex tail dependence structures, applied to us precipitation extremes. Journal of the American Statistical Association **115**, 1037–1054.

Castro-Camilo, D., Huser, R. and Rue, H. (2022) Practical strategies for generalized extreme value-based regression models for extremes. Environmetrics **33**, e2742.

Castruccio, S., Huser, R. and Genton, M. G. (2016) High-order composite likelihood inference for max-stable distributions and processes. Journal of Computational and Graphical Statistics **25**, 1212–1229.

Chandler, R. E. and Bate, S. (2007) Inference for clustered data using the independence loglikelihood. Biometrika **94**, 167–183.

Cotsakis, R., Di Bernardino, E. and Opitz, T. (2023) A local statistic for the spatial extent of extreme threshold exceedances. arXiv preprint arXiv:2310.09075 .

Davison, A. C. and Huser, R. (2015) Statistics of extremes. Annual Review of Statistics and its Application **2**, 203–235.

Davison, A. C., Huser, R. and Thibaud, E. (2019) Spatial extremes. In Handbook of Environmental and Ecological Statistics, eds A. E. Gelfand, M. Fuentes, J. A. Hoeting and R. L. Smith, pp. 711–744. CRC Press.

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

Dawkins, L. C. and Stephenson, D. B. (2018) Quantification of extremal dependence in spatial natural hazard footprints: independence of windstorm gust speeds and its impact on aggregate losses. Natural Hazards and Earth System Sciences **18**, 2933–2949.

Dombry, C., Engelke, S. and Oesting, M. (2016) Exact simulation of max-stable processes. Biometrika **103**, 303–317.

Dombry, C., Éyi-Minko, F. and Ribatet, M. (2013) Conditional simulation of max-stable processes. Biometrika **100**, 111–124.

Dombry, C. and Ribatet, M. (2015) Functional regular variations, pareto processes and peaks over threshold. Statistics and its Interface **8**, 9–17.

Donoho, D. (2017) 50 years of Data Science. Journal of Computational and Graphical Statistics **26**, 745–766.

Einmahl, J. H. J., Kiriliouk, A., Krajina, A. and Segers, J. (2016) An m-estimator of spatial tail dependence. Journal of the Royal Statistical Society: Series B **78**, 275–298.

Engelke, S. and Hitz, A. (2020) Graphical models for extremes (with discussion). Journal of the Royal Statistical Society: Series B **82**, 871–932.

Engelke, S., Opitz, T. and Wadsworth, J. (2019) Extremal dependence of random scale constructions. Extremes **22**, 623–666.

Fawcett, L. and Walshaw, D. (2007) Improved estimation for temporally clustered extremes. Environmetrics: The official journal of the International Environmetrics Society **18**, 173–188.

Ferreira, A. and de Haan, L. (2014) The generalized pareto process; with a view towards application and simulation. Bernoulli **20**, 1717–1737.

de Fondeville, R. and Davison, A. C. (2018) High-dimensional peaks-over-threshold inference. Biometrika **105**, 575–592.

de Fondeville, R. and Davison, A. C. (2022) Functional peaks-over-threshold analysis. Journal of the Royal Statistical Society Series B: Statistical Methodology **84**, 1392–1422.

Gabda, D., Towe, R., Wadsworth, J. and Tawn, J. A. (2012) Discussion of "statistical modeling of spatial extremes"' by a.c. davison, s.a. padoan and m. ribatet. Statistical Science **27**, 189–192.

Genton, M. G., Ma, Y. and Sang, H. (2011) On the likelihood function of gaussian max-stable processes. Biometrika **98**, 481–488.

de Haan, L. (1984) A spectral representation for max-stable processes. The Annals of Probability **12**, 1194–1204.

Hazra, A., Huser, R. and Bolin, D. (2022) Efficient modeling of spatial extremes over large geographical domains. arXiv preprint arXiv:2112.10248.

Hector, E. C. and Reich, B. J. (2023) Distributed inference for spatial extremes modeling in high dimensions. Journal of the American Statistical Association To appear.

Huser, R. and Davison, A. C. (2013) Composite likelihood estimation for the Brown–Resnick process. Biometrika **100**, 511–518.

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

Huser, R., Dombry, C., Ribatet, M. and Genton, M. G. (2019) Full likelihood inference for max-stable data. Stat **8**, e218.

Huser, R., Opitz, T. and Thibaud, E. (2017) Bridging asymptotic independence and dependence in spatial extremes using Gaussian scale mixtures. Spatial Statistics **21**, 166–186.

Huser, R., Opitz, T. and Thibaud, E. (2021) Max-infinitely divisible models and inference for spatial extremes. Scandinavian Journal of Statistics **48**, 321–348.

Huser, R., Stein, M. L. and Zhong, P. (2023) Vecchia likelihood approximation for accurate and fast inference in intractable spatial max-stable models. Journal of Computational and Graphical Statistics To appear.

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

Huser, R. and Wadsworth, J. L. (2022) Advances in statistical modeling of spatial extremes. Wiley Interdisciplinary Reviews: Computational Statistics **14**, e1537.

IPCC (2023) Climate Change 2023: Synthesis Report. In Contribution of Working Groups I, II, and III to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change, eds H. L. Core Writing Team and J. Romero, pp. 35–115. IPCC, Geneva, Switzerland.

Kabluchko, Z., Schlather, M. and de Haan, L. (2009) Stationary max-stable fields associated to negative definite functions. The Annals of Probability **37**, 2042–2065.

Krupskii, P. and Huser, R. (2022) Modeling spatial tail dependence with Cauchy convolution processes. Electronic Journal of Statistics **16**, 6135–6174.

Lafon, N., Naveau, P. and Fablet, R. (2023) A vae approach to sample multivariate extremes. arXiv preprint arXiv:2306.10987.

Ledford, A. W. and Tawn, J. A. (1996) Statistics for near independence in multivariate extreme values. Biometrika **83**, 169–187.

Ledford, A. W. and Tawn, J. A. (1997) Modelling dependence within joint tail regions. Journal of the Royal Statistical Society: Series B (Statistical Methodology) **59**, 475–499.

Lenzi, A., Bessac, J., Rudi, J. and Stein, M. L. (2023) Neural networks for parameter estimation in intractable models. Computational Statistics and Data Analysis **185**, 107762.

Lindgren, F., Rue, H. and Lindström, J. (2009) An explicit link between gaussian fields and gaussian markov random fields: the stochastic partial differential equation approach (with discussion). Journal of the Royal Statistical Society: Series B **73**, 423–498.

Murphy-Barltrop, C. J. R., Mackay, E. and Jonathan, P. (2024) Inference for multivariate extremes via a semi-parametric angular-radial model. arXiv preprint arXiv:2401.07259.

Naveau, P., Huser, R., Ribereau, P. and Hannart, A. (2016) Modeling jointly low, moderate and heavy rainfall intensities without a threshold selection. Water Resources Research **52**, 2753–2769.

Nguyen, M., Veraart, A. E. D., Taisne, B., Tan, C. T. and Lallemant, D. (2023) A dynamic extreme value model with application to volcanic eruption forecasting. Mathematical Geosciences to appear.

Nolde, N. (2014) Geometric interpretation of the residual dependence coefficient. Journal of Multivariate Analysis **123**, 85–95.

Nolde, N. and Wadsworth, J. L. (2022) Linking representations for multivariate extremes via a limit set. Advances in Applied Probability **54**, 688–717.

Oesting, M., Kabluchko, Z. and Schlather, M. (2012) Simulation of Brown–Resnick processes. Extremes **15**, 89–107.

Oesting, M. and Schlather, M. (2013) Conditional sampling for max-stable processes with a mixed moving maxima representation. Extremes **17**, 157–192.

Oesting, M. and Strokorb, K. (2022) A comparative tour through the simulation algorithms for max-stable processes. Statistical Science **37**, 42–63.

Opitz, T. (2013) Extremal t processes: Elliptical domain of attraction and a spectral representation. Journal of Multivariate Analysis **122**, 409–413.

Opitz, T. (2016) Modeling asymptotically independent spatial extremes based on Laplace random fields. Spatial Statistics **16**, 1–18.

Padoan, S. A., Ribatet, M. and Sisson, S. A. (2010) Likelihood-based inference for max-stable processes. Journal of the American Statistical Association **105**, 263–277.

Papastathopoulos, I., de Monte, L., Campbell, R. and Rue, H. (2023) Statistical inference for radially-stable generalized pareto distributions and return level-sets in geometric extremes. arXiv preprint arXiv:2310.06130.

Papastathopoulos, I. and Strokorb, K. (2016) Conditional independence among max-stable laws. Statistics and Probability Letters **108**, 9–15.

Radev, S. T., Mertens, U. K., Voss, A., Ardizzone, L. and Köthe, U. (2020) BayesFlow: Learning complex stochastic models with invertible neural networks. IEEE transactions on neural networks and learning systems **33**, 1452–1466.

Radev, S. T., Schmitt, M., Pratz, V., Picchini, U., Koethe, U. and Buerkner, P.-C. (2023a) JANA: Jointly amortized neural approximation of complex Bayesian models. In The 39th Conference on Uncertainty in Artificial Intelligence.

Radev, S. T., Schmitt, M., Schumacher, L., Elsemüller, L., Pratz, V., Schälte, Y., Köthe, U. and Bürkner, P.-C. (2023b) BayesFlow: Amortized Bayesian workflows with neural networks. arXiv preprint arXiv:2306.16015.

Reich, B. J. and Shaby, B. A. (2012) A hierarchical max-stable spatial model for extreme precipitation. The Annals of Applied Statistics **6**, 1430.

Resnick, S. (2002) Hidden regular variation, second order regular variation and asymptotic independence. Extremes **5**, 303–336.

Ribatet, M. (2022) SpatialExtremes: Modelling Spatial Extremes. R package version 2.1-0. https://CRAN.R-project.org/package=SpatialExtremes.

Richards, J., Sainsbury-Dale, M., Zammit-Mangion, A. and Huser, R. (2023) Likelihood-free neural bayes estimators for censored peaks-over-threshold models. arXiv preprint arXiv:2306.15642.

Risser, M. D. and Wehner, M. F. (2017) Attributable human-induced changes in the likelihood and magnitude of the observed extreme precipitation during Hurricane Harvey. Geophysical Research Letters **44**, 12457–12464.

Sainsbury-Dale, M., Richards, J., Zammit-Mangion, A. and Huser, R. (2023a) Neural Bayes estimators for irregular spatial data using graph neural networks. arXiv preprint arXiv:2310.02600.

Sainsbury-Dale, M., Zammit-Mangion, A. and Huser, R. (2023b) Likelihood-free parameter estimation with neural Bayes estimators. The American Statistician To appear.

Schlather, M. (2002) Models for stationary max-stable random fields. Extremes **5**, 33–44.

Segers, J. (2012) Max-stable models for multivariate extremes. REVSTAT **10**, 61–82.

Simpson, E. S., Opitz, T. and Wadsworth, J. L. (2023) High-dimensional modeling of spatial and spatio-temporal conditional extremes using INLA and Gaussian Markov random fields. Extremes pp. 1–45.

Simpson, E. S. and Wadsworth, J. L. (2021) Conditional modelling of spatio-temporal extremes for Red Sea surface temperatures. Spatial Statistics **41**, 100482.

Smith, R. L. (1990) Max-stable processes and spatial extremes. Unpublished manuscript. https://www.rls.sites.oasis.unc.edu/postscript/rs/spatex.pdf.

Stephenson, A. and Tawn, J. (2005) Exploiting occurrence times in likelihood inference for componentwise maxima. Biometrika **92**, 213–227.

Stott, P. A., Stone, D. A. and Allen, M. R. (2004) Human contribution to the european heatwave of 2003. Nature **432**, 610–614.

Thibaud, E. and Opitz, T. (2015) Efficient inference and simulation for elliptical Pareto processes. Biometrika **102**, 855–870.

Vandeskog, S., Martino, S. and Huser, R. (2022) An efficient workflow for modelling high-dimensional spatial extremes. arXiv preprint arXiv:2210.00760.

Wadsworth, J. and Campbell, R. (2022) Statistical inference for multivariate extremes via a geometric approach. arXiv preprint arXiv:2208.14951 .

Wadsworth, J. L. (2015) On the occurrence times of componentwise maxima and bias in likelihood inference for multivariate max-stable distributions. Biometrika **102**, 705–711.

Wadsworth, J. L. and Tawn, J. A. (2012) Dependence modelling for spatial extremes. Biometrika **99**, 253–272.

Wadsworth, J. L. and Tawn, J. A. (2022) Higher-dimensional spatial extremes via single-site conditioning. Spatial Statistics **51**, 100677.

Walchessen, J., Lenzi, A. and Kuusela, M. (2023) Neural likelihood surfaces for spatial processes with computationally intensive or intractable likelihoods. arXiv preprint arXiv:2305.04634.

White, H. (1982) Maximum likelihood estimation of misspecified models. Econometrica: Journal of the econometric society pp. 1–25.

Wikle, C. and Zammit-Mangion, A. (2023) Statistical deep learning for spatial and spatiotemporal data. Annual Review of Statistics and Its Application **10**, 247–270.

Yuen, R. and Stoev, S. (2014) CRPS M-estimation for max-stable models. Extremes **17**, 387–410.

Zhang, L., Ma, X., Wikle, C. K. and Huser, R. (2023a) Flexible and efficient spatial extremes emulation via variational autoencoders. arXiv preprint arXiv:2307.08079.

Zhang, Z., Bolin, D., Engelke, S. and Huser, R. (2023b) Extremal dependence of moving average processes driven by exponential-tailed Lévy noise. arXiv preprint arXiv:2307.15796.

Zhang, Z., Huser, R., Opitz, T. and Wadsworth, J. (2022) Modeling spatial extremes using normal mean-variance mixtures. Extremes **25**, 175–197.

Zhong, P., Huser, R. and Opitz, T. (2022) Modeling non-stationary temperature maxima based on extremal dependence changing with event magnitude. Annals of Applied Statistics **16**, 272–299.

Zscheischler, J., Martius, O., Westra, S., Bevacqua, E., Raymond, C., Horton, R. M. et al. (2020) A typology of compound weather and climate events. Nature Reviews Earth & Environment **1**, 333–347.