Modeling nonstationary extremes of storm severity: comparing parametric and semiparametric inference

Article

Published Version

It is advisable to refer to the publisher's version if you intend to cite from the work. See Guidance on citing.

www.reading.ac.uk/centaur

CentAUR

Central Archive at the University of Reading

Reading's research outputs online

WILEY

Modeling nonstationary extremes of storm severity: Comparing parametric and semiparametric inference

Evandro Konzen[1] | Cláudia Neves[1] | Philip Jonathan[2,3]

[1]Department of Mathematics and Statistics, University of Reading, Reading, UK

[2]Shell Research Ltd., London, UK

[3]Department of Mathematics and Statistics, Lancaster University, Lancaster, UK

Correspondence
Cláudia Neves, Department of Mathematics and Statistics, Whiteknights, PO Box 220, Reading RG6 6AX, UK.
Email: c.neves@reading.ac.uk

Funding information
Engineering and Physical Sciences Research Council, Grant/Award Number: EP/S001263/1; Project FCT, Grant/Award Number: UIDB/00006/2020

Abstract

This article compares the modeling of nonstationary extreme events using parametric models with local parametric and semiparametric approaches also motivated by extreme value theory. Specifically, three estimators are compared based on (a) (local) semiparametric moment estimation, (b) (local) maximum likelihood estimation, and (c) spline-based maximum likelihood estimation. Inference is performed in a sequential manner, highlighting the synergies between the different approaches to estimating extreme quantiles, including the T-year level and right endpoint when finite. We present a novel heuristic to estimate nonstationary extreme value threshold with exceedances varying on a circular domain, and hypothesis-testing procedures for identifying max-domain of attraction in the nonstationary setting. Bootstrapping is used to estimate nonstationary confidence bounds throughout. We provide step-by-step guides for estimation, and explore the different inference strategies in application to directional modeling of hindcast storm peak significant wave heights recorded in the North Sea.

KEYWORDS

circular, direction, endpoint, extreme quantile, kernel smoothing, peaks over threshold, periodic, significant wave height, spline, threshold selection

1 | INTRODUCTION AND MOTIVATION

Statistical inference for extreme values has developed rapidly in recent years, offering considerable scope for practical application in science and engineering. Two seemingly divergent camps of statistical thought have emerged, proposing different approaches to extreme value modeling, yielding inferences which are not always obviously in agreement. The work presented in this article seeks to identify potential points of contact between these so-called parametric and semiparametric frameworks for extreme value inference, to encourage better common understanding and convergence of at least some practices, in particular for tackling nonstationary extremes.

Nonstationarity is commonplace in environmental extremes; physical processes generate extreme values which typically vary systematically with covariates, including space and time. For the peaks-over-threshold (POT) method, where only data exceeding a threshold are used for analysis (Balkema & de Haan, 1974; Pickands III, 1975), various models have been proposed to capture nonstationarity, including those of Davison and Smith (1990) and Leadbetter (1991). In the parametric framework, nonstationarity can be incorporated within the appropriate distribution function for

threshold exceedances by allowing the distribution's (shape and scale) parameters to vary with covariates (see, e.g., Chavez-Demoulin & Davison, 2005; Coles, 2001, and references therein). Important assumptions underpinning this approach are that the data generating process is locally stationary and that observations from the data generating process can be considered approximately independent given covariates.

Applications of nonstationary extreme value analysis are more numerous using parametric inference. For example, in an ocean engineering context, Forristall (2004) performs extreme value analysis of significant wave height for directional octants. This approach (i) accommodates directional nonstationarity (by conducting independent inferences per directional octant) and (ii) allows extreme quantiles for specific directional sectors to be estimated. Choice of number and widths of directional sectors is an open problem (see, e.g., Folgueras et al., 2019; Ross et al., 2018). Moreover, nonstationary inference provides a number of challenges in practical assessment and interpretation of uncertainty and risk (e.g., Feld et al., 2019; Jonathan & Ewans, 2007; Mackay et al., 2010).

The need to incorporate a nonstationary threshold within extreme value inference is well recognized for environmental applications. Robinson and Tawn (1997) base their approach on a nonstationary threshold to characterize the evolution of extreme sea currents. Northrop and Jonathan (2011) propose a covariate-dependent threshold estimated using quantile regression, and Northrop et al. (2016) propose a cross-validation procedure for threshold selection. We introduce a method for selection of a nonstationary threshold amenable to both parametric and semiparametric approaches to inference. The basic assumption is that on a narrow region of the directional domain there is a unique physical regime which dictates the tail heaviness of the underlying distribution although extreme events can become more (or less) frequent. In this work, we propose a procedure for nonstationary threshold selection, which incorporates a smooth weight function to describe the varying density of extreme observations with directional covariate on the circular domain. The smooth weight function avoids inconsistencies stemming from the sectoring approach identified by Forristall (2004). The threshold selection procedure is common to both parametric and semiparametric approaches used for extreme value estimation.

Estimating the effects of periodic or circular covariates including direction and season in extreme value analysis is challenging in general. Coles and Walshaw (1994) consider extreme value analysis of wind speed on a circular domain, and Mendez et al. (2008) a seasonal model for significant wave height. Since extreme value parameters might be expected to vary smoothly with covariate, various different smooth basis representations (e.g., splines) for parameters of the conditional distribution of threshold exceedances, as a function of covariate as useful (see, e.g., Jones et al., 2016; Oumow et al., 2012; Ugarte et al., 2012; Zanini et al., 2020). Covariate basis representations require careful choice of order of complexity to avoid overparametrization. Some representations are attractive since basis functions have compact support on the covariate domain, facilitating computationally efficient and stable inference (Randell et al., 2016). In this work we consider a parametric estimation scheme using a penalized periodic B-spline covariate representation.

An alternative approach to nonstationary extreme value analysis is to estimate parametric and semiparametric models which are local on the covariate domain. Local models avoid the complexity of global covariate representations, particularly problematic for covariate intervals where extreme events are rare. In this work we consider semiparametric moment and local parametric maximum likelihood models.

Objectives and layout

The goal in this article is to explore the synergies between different parametric and semiparametric approaches to inference on extreme value data nonstationary with respect to periodic covariate. The default parametric approach employed is maximum likelihood estimation for peaks over threshold assumed to follow a generalized Pareto distribution, with spline descriptions of model parameter variation with covariate on some covariate domain. For comparison, two other covariate-local stationary models are considered: the first uses local parametric maximum likelihood estimation, and the second local semiparametric moment estimation; to achieve the latter we extend the semiparametric methodology so that inference for quantities of interest in ocean engineering is possible and meaningful. We develop (a) a new common local heuristic procedure for extreme value threshold selection for both local parametric and semiparametric frameworks, and (b) local inference for generalized Pareto shape parameter, extreme value index, and related quantities.

We then present a comparative study showing how the spline parametric, local parametric, and semiparametric approaches provide complementary inference for directional extremes in terms of (a) extreme value threshold selection, (b) extreme value model estimation, and (c) extreme quantile estimation including the T-year value and right endpoint of the support of the underlying distribution when is finite. We illustrate the approaches in application to directional extreme value analysis of hindcast storm peak significant wave height recorded at a northern North Sea location offshore

FIGURE 1 Left: map showing the North Sea location. Right: polar scatter-plot of H_S^{sp} on the direction θ from which waves propagate; radial scale of H_s^{sp} is in meters
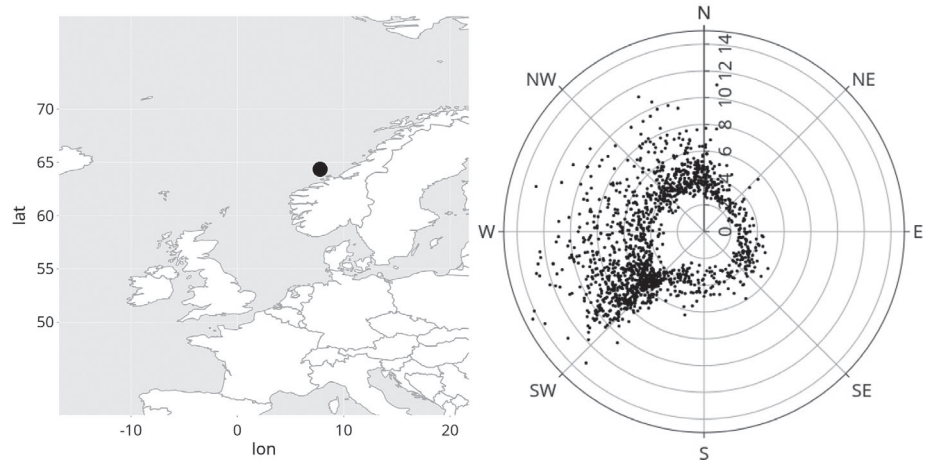
Norway. We conduct hypothesis tests to characterize variation in the max-domain of attraction of the distribution of storm peak significant wave height as a function of covariate. We hope to demonstrate that spline parametric, local parametric, and semiparametric approaches provide complementary nonstationary extreme value inference, and that local parametric and semiparametric approaches may be of practical benefit to practitioners in coastal and ocean engineering and environmental sciences.

The remainder of the article is organized as follows. The application to directional modeling of H_s^{sp} is introduced in Section 2. Section 3 provides theoretical motivation for nonstationarity analysis. Section 4 discusses nonstationary threshold estimation and extreme value model parameter estimation. Sections 5 and 6 discuss estimation of extreme value parameters and extreme quantiles. Results of the North Sea application are presented in Section 7. Finally, Section 8 provides discussion and conclusions. Appendix A provides theoretical motivation for nonstationary inference on peaks over threshold, and Appendix B provides algorithms for the estimation of extreme value threshold, spline generalized Pareto inference and for the overall analysis.

2 | MOTIVATING APPLICATION

The data for the motivating application are described by Randell et al. (2016), consisting of observations of storm severity, in terms of storm peak significant wave height henceforth referred to as H_s^{sp} and storm direction θ at a location in the northern North Sea. Significant wave height H_s measures the roughness of the ocean surface, and can be defined as four times the standard deviation of the ocean surface elevation at a spatial location for a specified period of observation. H_s^{sp} is then the maximum value of H_s observed per time period corresponding to a storm event. Values of H_s^{sp} can be regarded as approximately independent given storm peak characteristics.

The application sample is taken from the WAM hindcast of Reistad et al. (2011), which provides time-series of significant wave height, (dominant) wave direction and season (defined as day of the year, for a standardized year consisting of 360 days) for 3-h sea-states for the period September 1957 to December 2012 at a northern North Sea location in the vicinity of the black disk in the upper panel of Figure 1. A hindcast is a physical model of the ocean environment, incorporating pressure field, wind field and wind-wave generation models in particular; the hindcast model is calibrated to observations of the environment from instrumented offshore facilities, moored buoys, and satellite altimeters in the neighborhood of the location for a period of time, typically decades. Extreme seas in the North Sea are dominated by winter storms originating in the Atlantic Ocean and propagating eastward across the northern part of the North Sea. Due to their proximity to the storms, sea states at northern North Sea locations are usually more intense than in the southern North Sea. Occasionally, the storms travel south-eastward and intrude into the southern North Sea producing large sea states. Directions of propagation of extreme seas vary considerably with location, depending on land shadows of the British Isles, Scandinavia, and the coast of mainland Europe, and fetches associated with the Atlantic Ocean, Norwegian Sea, and the North Sea itself. In the northern North Sea the main fetches are the Norwegian Sea to the North, the Atlantic Ocean to the west, and the North Sea to the south. Extreme sea states from the directions of Scandinavia to the east and the British Isles to the south-west are unlikely. The shielding by these land masses is more effective for southern North Sea locations, resulting in a similar directional distribution but reduced wave heights by comparison with northern North Sea locations.

At the location of interest, observations of storm peak significant wave height H_s^{sp} are isolated from the hindcast time-series using the procedure described in Ewans and Jonathan (2008) as follows. Briefly, contiguous intervals of significant wave height above a low peak-picking threshold are identified, each interval assumed to correspond to a storm event. The peak-picking threshold corresponds to a directional-seasonal quantile of H_s with specified nonexceedance probability, estimated using quantile regression. The maximum of significant wave height during the storm interval is taken as the storm peak significant wave height H_s^{sp}. The values of directional and seasonal covariates at the time of storm peak significant wave height are referred to as storm peak values of those variables. The resulting storm peak sample consists of 2941 values. With direction *from which* a storm travels expressed in degrees clockwise with respect to north, the lower panel of Figure 1 shows a polar plot of observations of H_s^{sp} (in meters) versus direction θ.

The land shadow of Norway (approximately the directional interval (45°, 210°)) has a considerable effect on the rate and size of occurrences with direction. In particular, there is a dramatic increase in both rate and size of occurrences with increasing direction at around 210°, corresponding to Atlantic storm events from the south-west able to pass the Norwegian headland. We therefore expect considerable directional variability in extreme value parameter estimates for the sample. In contrast, the magnitude of the rate of change of both rate and size of occurrences with respect to season (not shown, but see Randell et al., 2016) is lower. Winter storms (approximately from October to March) are more intense and frequent. Only winter storms with storm peak events occurring in October to March including have been considered further in this work, corresponding to a sample of 1521 storm peak values.

3 | THEORETICAL MOTIVATION FOR THE NONSTATIONARY CASE

In Section 3.1 we first summarize theoretical results and assumptions underpinning extreme value modeling for the stationary (or omni-directional, or covariate-free) case. In Section 3.2 we then present a framework for local modeling of nonstationary extremes with covariate over a circular directional domain. These form the basis of inference schemes presented in Section 4.

3.1 | Stationary case

Suppose the available sample consists of realizations of n independent (or weakly dependent) and identically distributed random variables $X_1, X_2, \ldots, X_n$. Since the random variables follow the same (unknown) distribution function, for brevity use the symbol X to refer to any of the random variables when there is no need to be specific. The common distribution function is then $F(x) = P(X \leq x)$, for every $x \in \mathbb{R}$. We denote by x^F the right endpoint of the support of F, namely the ultimate value which bounds all possible observations from above, $x^F = \sup\{x : F(x) < 1\}$. We note that x^F may be less than or equal to infinity.

The extreme value (or extreme types) theorem (Fisher & Tippett, 1928; Gnedenko, 1943; de Haan, 1970) establishes that the limit distribution of linearly normalized partial maxima $M_n = \max(X_1, \ldots, X_n)$ (informally "block maxima"), with normalizing constants $a_n > 0$ and $b_n \in \mathbb{R}$, must be one of only possible three extreme value distributions: Fréchet, Gumbel, or Weibull. These three types can be nested in the one-parameter generalized extreme value (GEV) distribution. Specifically, if a_n and b_n exist such that,

$$\lim_{n \to \infty} P\left\{a_n^{-1}(M_n - b_n) \leq x\right\} = \lim_{n \to \infty} F^n(a_n x + b_n) = G(x), \tag{1}$$

for every continuity point of (nondegenerate) G, then G must be a GEV distribution with distribution function given by:

$$G_\xi(x) = \exp\left(-(1 + \xi\, x)^{-1/\xi}\right), \tag{2}$$

for all x such that $1 + \xi x > 0$. We then say that F belongs to the max-domain of attraction of the GEV, for some $\xi \in \mathbb{R}$, and write $F \in \mathcal{D}(G_\xi)$. Parameter ξ is conventionally referred to as the shape parameter in parametric literature, and the extreme value index (EVI) in semiparametric literature of extremes. A max-domain of attraction is thought of as a class enclosing all distribution functions which, in the limit, are associated with a particular extreme value index ξ. For example, both Cauchy and Student's t-distribution with one degree of freedom are found in the same max-domain of attraction $\mathcal{D}(G_\xi)$, with $\xi = 1$. The Fréchet max-domain of attraction, corresponding to $\xi > 0$, contains all distributions exhibiting

polynomial decay, such as the Pareto, again Cauchy and Student's t and the Fréchet itself. These distributions have infinite right endpoint x^F. All distribution functions belonging to $\mathcal{D}(G_\xi)$ with $\xi < 0$, referred to as the Weibull max-domain of attraction, are short-tailed with finite x^F; examples include the uniform and beta distribution. For $\xi = 0$, a continuity argument gives $G_{\xi=0}(x) = \exp(-e^{-x})$, $x \in \mathbb{R}$; the corresponding Gumbel max-domain of attraction ($\mathcal{D}(G_\xi)$ with $\xi = 0$, or $\mathcal{D}(G_0)$), of particular interest in many applied fields due to both simplicity of inference and the great variety of distributions possessing an exponential tail (with either $x^F < \infty$ or $x^F = \infty$). The normal, gamma and log-normal distributions are members of $\mathcal{D}(G_0)$.

Extreme value inference using block maxima has been practiced for many years, from the time of Gumbel (1958) in application to hydrology. However, in this article, inference is based on analysis of threshold exceedances or peaks over threshold (POT), motivated by characterization of the max-domains of attraction above in terms of exceedances of high threshold. Balkema and de Haan (1974) and Pickands III (1975) established that the max-domain of attraction condition (1) is equivalent to the assertion that the (conditional) distribution of X given $X > u$, with u near the right endpoint x^F, converges to the generalized Pareto distribution (GPD) with distribution function $1 + \log G_\xi$. Analysis of threshold exceedances is potentially statistically more efficient than that of block maxima: in the former, all large values of threshold exceedance in the sample are admitted, including multiple occurrences of large values belonging to same block, which might be excluded in a block maximum analysis.

The parametric and semiparametric approaches to extreme value analysis of threshold exceedances considered in this article have strong conceptual similarities, and both approaches involve estimation of three quantities corresponding to threshold, generalized Pareto shape and scale in parametric inference, and threshold, extreme value index (EVI) and associated scale in semiparametric inference. There is a large literature on estimation of shape parameter in parametric inference, and the extreme value index in semiparametric inference. In this work we use maximum likelihood (ML) (Coles, 2001; Davison & Smith, 1990; Smith, 1987) and moment (M) estimators (Dekkers et al., 1989) as typical parametric and semiparametric estimators, due to their popularity. The ML estimator assumes that the limiting generalized Pareto distribution provides an accurate description of the largest values in the sample, as though the exceedances over a fixed threshold were drawn from an actual generalized Pareto distribution; this provides a highly flexible framework for inference. The moment estimator builds on the concept of max-domain of attraction, and in this sense is model-free; it is more robust: there are seldom any assurances that exceedances of high threshold actually follow the generalized Pareto form. The moment estimator can be viewed as intuitive, since its form involves summary statistics like averages. When goodness of fit can assessed with reasonable accuracy, the ML estimator can be shown to be efficient for the shape parameter. There is concern however that the maximum likelihood may be influenced by deviations of the true underlying distribution function from the assumed generalized Pareto, resulting in extrapolation bias. For this reason, it is sometimes preferable to consider a robust estimator building on moments to lessen the importance of such deviations, at the expense of increased variance.

Viewed parametrically or semiparametrically, choosing an appropriate extreme value threshold is fundamental to inference for peaks over threshold, and involves a typical bias-variance trade-off. If the threshold is set too high, the set of threshold exceedances is small resulting in increased variance of parameter estimates. If the threshold is set too low, the variance of the parameter estimate is reduced at the cost of increased bias. Sensitivity to the extreme value threshold choice is a common critical feature of parametric and semiparametric approaches. Confirming the relative stability of estimated shape parameter (or extreme value index) and high quantiles (such as the T-year level) with near-zero exceedance probability with respect to threshold is a key diagnostic test for peaks over threshold analysis, addressed further in Section 6.

Given threshold, inference for peaks over threshold in the parametric setting involves estimation of generalized Pareto shape and scale parameters, the latter being threshold-dependent; for estimation of T-year levels, we also need to estimate the rate of threshold exceedance. In the semiparametric setting, we focus on estimation of extreme value index; scale and location normalizing functions (akin to $a > 0$ and $b \in \mathbb{R}$ in (A5)) required for extreme quantile calculations, are estimated separately.

3.2 | Nonstationary case

In the nonstationary setting, we assume that the covariate domain S is a circle, partitioned into m intervals, with centroids θ_j, and write Θ for the set $\{\theta_j\}_{j=1}^m$ of centroids; for directional analysis considered here, we set $S = [0°, 360°)$, and $\theta_j = j - 1$ for $j = 1, \ldots, m = 360$. For each θ_j, we suppose there are n independent copies $\{X_i(\theta_j)\}_{i=1}^n$ of random variable

$X(\theta_j)$ with distribution function $F_{\theta_j}(x) = P\{X(\theta_j) \le x\}, x \in \mathbb{R}$. We also assume that $X(\theta_j)$ and $X(\theta_{j'})$ are independently distributed when $j \neq j'$. We then assume the extreme value theorem holds on covariate neighborhood $\mathcal{N}(\theta_j, h)$ of half-width $h > 0$ centered at θ_j (see Section 4.1). This means that condition (1) holds with respect to F_{θ_j} and the shape parameter (or extreme value index) $\xi(\theta_j)$ governs the tail behavior of the underlying F_{θ_j}. We also assume throughout that changes in the tail dynamics of the data generating process are smooth, and that $\xi(\theta)$ is a smooth (unknown) function of $\theta \in \mathcal{S}$. These assumptions justify the nonstationary parametric and semiparametric inference procedures used in this work.

The theoretical motivation for the nonstationary case is given in Appendix A, and summarized here. Suppose that for each $\theta \in \mathcal{S}$, neighborhood $\mathcal{N}(\theta, h)$ contains $N(\theta)$ random occurrences of X. We start by selecting $k(\theta)$ threshold exceedances on that neighborhood, assuming a limiting generalized Pareto distribution holds for sufficiently large $N(\theta)$. Given the initial assumption of (weak) dependence on $\mathcal{S}$, tail observations from F_θ can be viewed as peaks over threshold at each θ as follows, for any $\theta \in \mathcal{N}(\theta, h) \subset \mathcal{S}$. With the notation $N = n \times m$, let $X_{N-k(\theta):N}$ denote the $(N - k(\theta))$th neighborhood-specific threshold having selected fixed number $k(\theta)$ for that neighborhood. To ensure integrity of statistical analysis on $\mathcal{S}$, the number $k(\theta)$ is defined using a weighting function $\omega(\theta_j) > 0$ so that $k(\theta) := [k \times \omega(\theta)]$, with $[\cdot]$ indicating "integer part of," and k is the total number of largest observations retained for inference across the whole of $\mathcal{S}$. For a finite number m of directional sectors θ_j, the smooth function $\omega(\theta)$ satisfies $\sum_{j=1}^m \omega(\theta_j)/m = 1$.

Given threshold $u(\theta) = X_{N-k(\theta):N}$, we perform nonstationary extreme value inference for peaks over threshold. In the parametric setting, inference involves estimation of generalized Pareto shape and scale parameters which vary on the covariate domain; for estimation of T-year levels, we also need to estimate the rate of threshold exceedance which also varies with covariate. The semiparametric approach is concerned with extrapolation beyond the sample through estimation of extreme quantiles with very small tail probabilities; semiparametric estimators are typically smooth functionals (averages, quantiles, etc.) of sample upper order statistics, estimating extreme value index and scale function, in terms of the extreme sample fraction $k(\theta)/N$.

4 | NONSTATIONARY THRESHOLD SELECTION

Here we describe an approach to nonstationary threshold selection useful for all of parametric and semiparametric settings. The approach exploits a heuristic criterion to select a number $k = k(\theta)$ of upper order statistics, and hence a random threshold on the covariate domain. Algorithm 1 of Appendix B provides a simple implementation of the threshold selection procedure. In Section 5 we then discuss how the estimated threshold is used in extreme value estimation using each of the parametric, local parametric, and semiparametric approaches. Note that although threshold estimation is made here using either the semiparametric moment estimator (Section 4.2) or ML estimation (Section 4.3), any other consistent estimator would be equally applicable.

4.1 | Threshold selection

Let $\mathcal{N}(\theta_j, h)$ be the covariate neighborhood with center at $\theta_j \in \Theta$ and fixed radius $h > 0$ defined by

$$\mathcal{N}(\theta_j, h) = \{\theta^* \in \mathcal{S} : \ 0 \le d(\theta^*, \theta_j) \le h\}, \tag{3}$$

for every θ_j for an appropriate choice of distance d: in the directional application, we adopt wrapped-Euclidean distance $d(\theta^*, \theta_j) := \min\{|\theta^* - \theta_j|, 360 - |\theta^* - \theta_j|\}$ for $\theta^* \in \mathcal{S} = [0,360)$. For each θ_j, the number $N(\theta_j)$ of realizations of $X_i(\theta_{j'})$, $i = 1, \ldots, n$, $\theta_{j'} \in \Theta$, in the neighborhood within lag h of centroid θ_j is

$$N(\theta_j) = |C(\theta_j, h)|$$

and $C(\theta_j, h) = \{X_i(\theta_{j'}) \mid \theta_{j'} \in \mathcal{N}(\theta_j, h), i = 1, 2, \ldots, n\}$ A judicious choice of h ensures a large enough number $N(\theta_j)$ is present, so that the extreme value theorem holds on $\mathcal{N}(\theta_j, h)$, for every θ_j.

We now consider a heuristic threshold selection procedure exploiting the relative density of extreme values in $\mathcal{N}(\theta_j, h)$. We seek the optimal number k_j^* of the k largest observations in $\mathcal{N}(\theta_j, h)$ as threshold exceedances. Then k_j^* is the solution of $S_\phi(k_j^*) = \min_k S_\phi(k)$ with

$$S_\phi(k) = \frac{1}{k}\sum_{i\le k} i^\phi \left|\hat\xi_i(\theta_j) - \text{median}\left(\hat\xi_1(\theta_j), \hat\xi_2(\theta_j), \ldots, \hat\xi_k(\theta_j)\right)\right|, \tag{4}$$

where $0 \le \phi < 0.5$, and $\hat\xi_k(\theta_j)$ is a locally consistent estimator of $\xi(\theta_j)$ using the k largest observations in $\mathcal{N}(\theta_j, h)$. The heuristic procedure (4) was introduced in the first edition of Reiss and Thomas (the most recent edition being Reiss & Thomas, 2007) and then studied in detail in Neves and Fraga Alves (2004). It facilitates automatic threshold choice, understood intuitively as follows. For small k, the weighted deviations in (4) tend to be large due to the inherently large variance of $\hat\xi_k(\theta_j)$. As k increases, the summands in (4) are expected to decrease until bias sets in and overrides the variance from which point $S_\phi(k)$ is expected to increase again. Minimizing the weighted empirical distance (4) is equivalent to optimizing the bias-variance trade-off by exploiting the settled behavior of estimates $\{\hat\xi_k(\theta_j) :\ k < N(\theta_j)\}$ for appropriate k.

We specify a covariate weight $\omega(\theta_{j'}; \theta_j)$ of observations for every $\theta_{j'} \in \mathcal{N}(\theta_j, h) \cap \Theta$ relative to centroid $\theta_j \in \Theta$ using a von Mises kernel (cf., Pewsey et al., 2014), such that

$$\omega(\theta_{j'}; \theta_j) := \frac{\mathcal{K}_\eta(\theta_{j'}; \theta_j)}{\sum_{\theta_{j'} \in \mathcal{N}(\theta_j, h)} \mathcal{K}_\eta(\theta_{j'}; \theta_j)}, \tag{5}$$

with von Mises kernel $\mathcal{K}_\eta(\theta; \theta_j) := \exp\{\eta\cos(\theta - \theta_j)\}/(2\pi B_0(\eta))$ for $\theta \in \mathcal{S}$, where B_0 is the modified Bessel function of the first kind of order zero $B_0(\eta) = \pi^{-1}\int_0^\pi \exp[\eta\cos(\theta)]d\theta$. Concentration parameter $\eta > 0$ controls kernel spread: as η increases, the local density becomes more concentrated about θ_j; thus, η plays a role similar to bandwidth h in (3), regulating degree of directional smoothness. The values of η and h need to specified carefully during applications to reflect the characteristics of the data generating process, and the resulting choice of k_j^* reflects the extent of stationarity on $\mathcal{N}(\theta_j, h)$. The corresponding optimal estimate for nonstationary threshold is denoted $u(\theta_j)$ for each $\theta_j \in \Theta$.

In this section we indicate the dependence of $\omega(\theta_{j'}; \theta_j)$ on centroid θ_j explicitly for clarity. In the discussion in Appendix A, we write $\omega(\theta_{j'})$ for notational convenience, with the dependence on θ_j understood. We do not claim that this heuristic approach to extreme value threshold specification is the most effective (e.g., with respect to every estimator one might devise for extreme value index), but we have found it useful in the application to estimation of ξ in Section 7.

4.2 | Semiparametric estimation

Now we outline semiparametric estimation of $\xi_k(\theta_j)$ required for threshold selection above. For each $\theta_j \in \Theta$, we propose an adaption of the extended version of the moment estimator (Dekkers et al., 1989) for $\xi(\theta_j)$, which incorporates directional weights $\omega(\theta_{j'}; \theta_j)$ defined in (5). Let $h > 0$ be a bandwidth such that $h = h(N(\theta_j)) \to 0$ and $k(\theta)h \to \infty$, as $N(\theta_j) \to \infty$. The moment estimator for the extreme value index is defined as

$$\hat\xi_k^M(\theta_j) := M^{(1)}(\theta_j) + 1 - \frac{1}{2}\left(1 - \frac{\left(M^{(1)}(\theta_j)\right)^2}{M^{(2)}(\theta_j)}\right)^{-1}, \tag{6}$$

with

$$M^{(l)}(\theta_j) := \sum_{C(\theta_j, h)} \omega(\theta_{j'}; \theta_j)\left(\log X_i(\theta_{j'}) - \log X_{N(\theta_j) - k, N(\theta_j)}\right)^l \mathbb{1}_{\{X_i(\theta_{j'}) > X_{N(\theta_j) - k, N(\theta_j)}\}}, \quad l = 1, 2, \tag{7}$$

where $X_{N(\theta_j) - k, N(\theta_j)}$ denotes the $(k+1)$th largest value in the observed sample in $\mathcal{N}(\theta_j, h)$. Note only the largest k observations in $\mathcal{N}(\theta_j, h)$ contribute in (7). Further discussion of the moment estimator is given in Section 5.1.

4.3 | Maximum likelihood estimation

The local ML estimator $\hat\xi_k(\theta_j)$ for $\theta_j \in \Theta$ for threshold selection above is obtained by maximizing the directionally weighted log-likelihood

$$\mathcal{L}\left(\xi(\theta_j), \sigma_u(\theta_j)\right) := \sum_{C(\theta_j, h)} \omega(\theta_{j'}; \theta_j)\,\ell\left(\xi(\theta_j), \sigma_u(\theta_j)|\,X_i(\theta_{j'}) - X_{N(\theta_j) - k, N(\theta_j)}\right)\mathbb{1}_{\{X_i(\theta_{j'}) - X_{N(\theta_j) - k, N(\theta_j)} > 0\}}, \tag{8}$$

with respect to the parameter pair $\left(\xi(\theta_j), \sigma_u(\theta_j)\right) \in (-1, \infty) \times \mathbb{R}^+$, with weights $\omega(\theta_{j'}; \theta_j)$ as in Section 4.2, and

$$\ell\left(\xi(\theta_j), \sigma_u(\theta_j)\mid y\right) = -\log \sigma_u(\theta_j) - \left(1 + 1/\xi(\theta_j)\right) \log\left(1 + \xi(\theta_j)\, y/\sigma_u(\theta_j)\right) \tag{9}$$

for $y = X_i(\theta_{j'}) - X_{N(\theta_j)\,-k,N(\theta_j)}$ when $\xi(\theta_j) \neq 0$ and $\ell\left(\sigma_u(\theta_j)\mid y\right) = -\log \sigma_u(\theta_j) - y/\sigma(\theta_j)$ otherwise.

5 | NONSTATIONARY EXTREME VALUE ESTIMATION

This section considers estimation of extreme value parameters and extreme quantiles. We first present adapted versions of widely used semiparametric estimators for extreme value index and associated scale function based on a known moving threshold $u(\theta), \theta \in S$. Then we discuss two classes of parametric maximum likelihood estimators for generalized Pareto shape and scale. The first is a minor modification of the local ML estimator introduced in Section 4.3. The second is motivated by a spline representation of the variation of generalized Pareto shape and scale with respect to covariate. Section 6 then addresses estimation of extreme quantiles using the estimated parameters.

In this and subsequent sections, we consider the nonstationary extreme value threshold $u(\theta)$ to be determined and fixed; our focus is therefore on characterizing threshold exceedances. Regardless of the approach to extreme value estimation, the heuristic procedure of Section 4 is used to estimate $u(\theta)$.

Descriptions of the semiparametric, local parametric, and spline parametric estimation schemes are given in Sections 5.1–5.3.

5.1 | Local semiparametric moment estimation

Given threshold $u(\theta_j)$ for $\theta_j \in \Theta$, we adopt the moment estimator for the extreme value index given by

$$\hat{\xi}_k^M(\theta_j) := M^{(1)}(\theta_j) + 1 - \frac{1}{2}\left(1 - \frac{\left(M^{(1)}(\theta_j)\right)^2}{M^{(2)}(\theta_j)}\right)^{-1}, \tag{10}$$

where moments $M^{(l)}(\theta_j)$, $l = 1, 2$ are calculated using

$$M^{(l)}(\theta_j) = \frac{1}{k(\theta_j)} \sum_{C(\theta_j,h)} \left(\log X_i(\theta_{j'}) - \log u(\theta_j)\right)^l \mathbb{1}_{\{X_i(\theta_{j'}) > u(\theta_j)\}}, \tag{11}$$

where $k(\theta_j)$ is the number of threshold exceedances in $\mathcal{N}(\theta_j, h)$. Equation (11) corresponds to an unweighted version of the moment estimator used for threshold estimation in Equation (7).

This moment estimator is motivated by the asymptotic behavior of the $k(\theta_j)$th upper order statistics associated with the sample of independent random variables $X_i(\theta_{j'}) - X_{N(\theta_j)\,-k(\theta_j),N(\theta_j)}$ on $\mathcal{N}(\theta_j, h)$ as established in the theory of extremes for threshold excesses. Conditionally on $X_{N(\theta_j)\,-k(\theta_j),N(\theta_j)} = u(\theta_j)$, their underlying distribution function tends to $F^{[u(\theta_j)]}(t) = P\left(X(\theta_j) - u(\theta_j) > t\mid X(\theta_j) > u(\theta_j)\right)$, for all $t > 0$ (cf., de Haan & Ferreira, 2006, p. 90). Smith (1987) defined the analogous ML estimator in terms of excesses over a high threshold u. We refer the reader to Appendix A for a precise definition of $F^{[u]}$, and further details on how it approaches the GPD distribution function.

An associated scale estimator (e.g., de Haan & Ferreira, 2006, section 4.2) necessary for semiparametric estimation of extreme quantiles in Section 6 is then given by

$$\hat{a}_{\theta_j}^M\left(\frac{N}{k(\theta_j)}\right) := u(\theta_j)\, M^{(1)}(\theta_j) \frac{1}{2}\left(1 - \frac{(M^{(1)}(\theta_j))^2}{M^{(2)}(\theta_j)}\right)^{-1}. \tag{12}$$

The moment estimator for the scale function $a_{\theta_j}^*\left(N/k(\theta_j)\right)$ is discussed in Appendix A(ii).

5.2 | Local maximum likelihood estimation

Given threshold $u(\theta_j)$ for $\theta_j \in \Theta$, local ML estimation described of generalized Pareto shape $\xi(\theta_j)$ and scale $\sigma_u(\theta_j)$ is performed by the procedure described in Section 4.3, with all directional weights ω set to unity. For completeness, the log-likelihood to be maximized is given by

$$\mathcal{L}\left(\xi(\theta_j), \sigma_u(\theta_j)\right) := \sum_{C(\theta_j, h)} \ell\left(\xi(\theta_j), \sigma_u(\theta_j) | X_i(\theta_{j'}) - u(\theta_j)\right) \mathbb{1}_{\{X_i(\theta_{j'}) - u(\theta_j) > 0\}}. \tag{13}$$

This formulation incorporates a semiparametric approach to threshold selection within a local parametric approach to parameter estimation: conditioned on the random number $K = k(\theta_j)$ of exceedances, random exceedances of threshold $u(\theta_j)$ can again be viewed as independent identically distributed random variables with distribution function converging asymptotically to the generalized Pareto (Appendix A).

5.3 | Spline-based maximum likelihood estimation

Given nonstationary threshold $u(\theta_j)$ for $\theta_j \in \Theta$, we can also estimate a nonstationary generalized Pareto model using the full sample of threshold exceedances, by allowing generalized Pareto shape ξ and scale σ_u to be functions of θ on S. This is the type of analysis typically undertaken in environmental and ocean engineering settings (see, e.g., Northrop et al., 2016). Specifically, we assume a periodic cubic B-spline representation for the variation of generalized Pareto shape and scale parameters with covariate (see, e.g., chapter 5 of Wood, 2017; Zanini et al., 2020). We then estimate spline coefficients using maximum penalized likelihood estimation, regulating the roughness of shape and scale with covariate to optimize predictive performance assessed using cross-validation.

On the index set Θ of covariate values θ_j, $j = 1, 2, \ldots, m$, we relate the values $\xi(\theta_j), \sigma_u(\theta_j)$ of generalized Pareto shape and scale to the periodic B-spline basis via basis matrix $\boldsymbol{B}$ with elements B_{jb} such that

$$\xi(\theta_j) = \sum_{b=1}^{n_b} B_{jb} \beta_b^{(1)} \quad \text{and} \quad \log \sigma_u(\theta_j) = \sum_{b=1}^{n_b} B_{jb} \beta_b^{(2)}, \quad \theta_j \in \Theta, \tag{14}$$

where n_b is the number of basis functions, and the βs are basis coefficients. The sample log-likelihood $\mathcal{L}$ is

$$\mathcal{L}(\boldsymbol{\beta}) = \sum_{j=1}^{m} \sum_{i=1}^{n(\theta_j)} \ell\left(\xi(\theta_j), \sigma_u(\theta_j) | X_i(\theta_j) - u(\theta_j)\right) \mathbb{1}_{\{X_i(\theta_j) > u(\theta_j)\}}, \tag{15}$$

with $\ell(\xi(\theta_j), \sigma_u(\theta_j) | X_i(\theta_j) - u(\theta_j))$ defined in (9). $n(\theta_j)$ denotes the number of observations in the sample at covariate θ_j, $\boldsymbol{\beta}^{(a)} = (\beta_1^{(a)}, \beta_2^{(a)}, \ldots, \beta_{n_b}^{(a)})^\top$ for $a = 1, 2$, and $\boldsymbol{\beta} = (\boldsymbol{\beta}^{(1)\top}, \boldsymbol{\beta}^{(2)\top})^\top$. We set the number of spline knots on S to be more than sufficient to capture the anticipated parameter variability with covariate, and then penalize parameter roughness globally to obtain a model with good predictive performance. Penalization is performed using first-order difference penalties for the coefficients in (14),

$$P^{(a)} = \sum_{b=1}^{n_b-1} \left(\beta_{b+1}^{(a)} - \beta_b^{(a)}\right)^2 + \left(\beta_1^{(a)} - \beta_n^{(a)}\right)^2 = \boldsymbol{\beta}^{(a)\top} \boldsymbol{D}^\top \boldsymbol{D} \boldsymbol{\beta}^{(a)}, \quad a = 1, 2,$$

with difference matrix $\boldsymbol{D}$ given by:

$$\boldsymbol{D} = \begin{bmatrix} -1 & 1 & 0 & 0 & \cdot & \cdot \\ 0 & -1 & 1 & 0 & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ 0 & 0 & \cdot & 0 & -1 & 1 \\ 1 & 0 & \cdot & 0 & 0 & -1 \end{bmatrix}.$$

The penalized log-likelihood is then:

$$\mathcal{L}_{pen}(\boldsymbol{\beta}) = \mathcal{L}(\boldsymbol{\beta}) - \lambda P^{(1)} - \kappa P^{(2)}, \tag{16}$$

where λ and κ are smoothing parameters chosen to maximize cross-validated predictive likelihood. In the application illustrated in Section 7, cross-validation is applied as follows. Each iteration of the cross-validation consists in using a bootstrap resample (of the original sample of threshold exceedances) as training set, and observations omitted from the bootstrap resample as test set. Note that the test sets corresponding to different bootstrap resamples may therefore overlap. The training set is used to estimate the model parameters for a grid of combinations of λ and κ, and the test set to assess prediction performance using predictive log likelihood. The choices of λ and κ provide best predicted performance are adopted for subsequent inferences. The procedure is outlined in Algorithm 2 in Appendix B.

6 | ESTIMATION OF NONSTATIONARY EXTREME QUANTILE

In this section we use extreme value estimates from Section 5 to calculate extreme quantiles, including the right endpoint of the distribution of $X(\theta)$ when this is finite, in the semiparametric setting (Section 6.1) and parametric setting (Section 6.2). The stepwise inference procedure is given in Algorithm 3 of Appendix B.

A common motivation applies for estimation of extreme quantiles in both parametric and semiparametric settings. Suppose that F_θ is the actual distribution function of the data generating process at direction $\theta \in S$ and Q_θ the corresponding quantile function; that is, $Q_\theta = F_\theta^\leftarrow$, with the left arrow indicating a generalized inverse. Conditions (A2) and (A3) of Appendix A ensure that an extreme quantile $F_\theta^\leftarrow(1-p)$, with $1 - p > F_\theta(u(\theta))$ for some high threshold $u(\theta)$ and hence very small probability p, depends only on the tail of the distribution function F_θ. Consequently, we can estimate an extreme quantile using linear functional $\hat{b}_\theta + \hat{a}_\theta \, Q_\theta^H(1-p)$, with Q_θ^H relating to a generalized Pareto distribution function H_θ. This result applies for both parametric and semiparametric approaches, with small differences. The most obvious difference is that in the parametric approach, normalizing constants a_θ and b_θ are provided by the scale and location of the limiting generalized Pareto distribution, whereas in the semiparametric approach these are estimated as functionals of the sample analogue of distribution function F_θ.

6.1 | Semiparametric-based estimation

6.1.1 | Extreme quantile

In the semiparametric setting, we assume that the number $k(\theta)$ of exceedances of threshold $u(\theta)$ is such that $k(\theta) \to \infty$ and $k(\theta)/N(\theta) \to 0$ as $N(\theta) \to \infty$. That is, the number of threshold exceedances $k(\theta)$ remains negligible compared with the total number $N(\theta)$ of observations in $\mathcal{N}(\theta, h)$. This condition can be written in terms of the covariate-specific sample size n, requiring that $n \to \infty$ implies $N(\theta) \to \infty$. The proposed moment estimator for the extreme quantile with small probability $p \to 0$, $N(\theta)p/k(\theta) \to 0$ as $n \to \infty$, conditioned on threshold $u(\theta)$, is

$$\hat{x}_p^M(\theta) = u(\theta) + \hat{a}_\theta^M \left(\frac{N}{k(\theta)}\right) \frac{\left(\frac{k(\theta)}{N(\theta)p}\right)^{\hat{\xi}_k^M(\theta)} - 1}{\hat{\xi}_k^M(\theta)}, \tag{17}$$

with estimated extreme value index $\hat{\xi}_k^M(\theta)$ and associated scale parameter $\hat{a}_\theta^M \left(\frac{N}{k(\theta)}\right)$ given in Section 5.1. The associate scale in (17) is a function of the top sample fraction $k(\theta)/N(\theta)$.

6.1.2 | Finite right endpoint

Semiparametric estimators for the right endpoint have received considerable attention in the literature (e.g., chapter 4 of de Haan & Ferreira, 2006, and connected references). We propose a model-free, data-driven estimator of the covariate-dependent right endpoint $x^{F_\theta} < \infty$ for $\theta \in S$, motivated by the general endpoint estimator of Fraga Alves and

Neves (2014) with nonstationary threshold $u(\theta)$. We proceed on the basis of a suitable extreme value condition which partitions the Gumbel max-domain of attraction (with extreme value index $\xi = 0$) into two classes of distributions, one with finite right endpoint, and the other right unbounded. An example of a distribution function in the former class is the negative Fréchet, with distribution function $F_{\alpha,\beta}(x) = 1 - \exp\{-(\alpha - x)^{-\beta}\}$, $x \leq \alpha$, $\alpha \in \mathbb{R}$, $\beta > 0$; simple calculations show that $F_{\alpha,\beta}$ belongs to $\mathcal{D}(G_0)$ despite having finite right endpoint $x^F = \alpha < \infty$. We seek a right endpoint estimator for distributions in this $\mathcal{D}(G_0)$ subclass.

With $\{Y_i\}_{i=1}^{N(\theta)}$ representing the random variables X_i in a particular neighborhood $\mathcal{N}(\theta, h)$, defined in (3), we denote by $Y_{1,N(\theta)} \leq \ldots \leq Y_{N(\theta)-k(\theta),N(\theta)} \leq \ldots \leq Y_{N(\theta),N(\theta)}$ the corresponding ascending order statistics, and define the general endpoint estimator of x^{F_θ}, assumed finite, for every $\theta \in S$,

$$\hat{x}_0^G(\theta) := Y_{N(\theta),N(\theta)} + \frac{1}{\log 2} \sum_{i=0}^{k(\theta)-1} \log\left(\frac{k(\theta) + i + 1}{k(\theta) + i}\right) \left(Y_{N(\theta)-k(\theta),N(\theta)} - Y_{N(\theta)-k(\theta)-i,N(\theta)}\right). \tag{18}$$

This estimator is valid for any $\xi(\theta) \leq 0$, and does not require prior estimation of extreme value index (cf., Fraga Alves et al., 2017), in contrast to competitor estimators for the right endpoint proposed in the literature.

6.2 | Parametric-based estimation

6.2.1 | Extreme quantile

In the parametric setting, level $x_p(\theta)$ with small exceedance probability p corresponds to the quantile of the distribution of threshold exceedances for direction θ. This suggests defining the $1/p$ extreme quantile as the value $x_p(\theta) = Q_\theta^H(1 - p)$, with normalizing constants set to the scale and location parameters of a generalized Pareto function H_θ. Hence, for local parametric and spline parametric approaches, estimates of $x_p(\theta)$ of the form

$$\hat{x}_p^{\mathrm{ML}}(\theta) = u(\theta) + \hat{\sigma}_u(\theta) \frac{\left(\frac{\varphi(u,\theta)}{p}\right)^{\hat{\xi}(\theta)} - 1}{\hat{\xi}(\theta)}, \tag{19}$$

are obtained, where $\varphi(u, \theta) = k(\theta)/N(\theta)$ is the sample fraction of exceedances of $u(\theta)$ within $\mathcal{N}(\theta, h)$. Estimation of $\hat{\xi}(\theta)$ and $\hat{\sigma}_u(\theta)$ (for $\theta \in \Theta$) using local parametric and spline parametric approaches is discussed in Sections 5.2 and 5.3. The extreme-level estimate $\hat{x}_p^{\mathrm{ML}}(\theta)$ can be interpreted as a T-year level used in ocean engineering and hydrology: that is, the value exceeded with probability $1/T$ per annum. Expressions (19) and (17) show obvious similarities, and also distinctive traits of the semiparametric moment, local parametric and spline parametric approaches.

For local ML estimation (Section 5.2), the values of $k(\theta_j)$ and $N(\theta_j)$ are estimated directly in Section 4. In the spline maximum likelihood approach (Section 5.3), a smoothed version of $\varphi(u, \theta_j)$ is estimated as the probability of threshold exceedance for $\theta_j \in \Theta$ using logistic regression, with log-likelihood

$$\mathcal{L}_{\mathrm{Lgs}}(\boldsymbol{\beta}) = \sum_{j=1}^{m} \tau_j \log[v_j] + (1 - \tau_j) \log[1 - v_j], \tag{20}$$

where $\tau_j = k(\theta_j)/N(\theta_j)$ is the sample proportion of threshold exceedances of $u(\theta_j)$ at $\theta_j \in \Theta$. $v_j = (1 + \exp[-\eta_j])$ is the probability of threshold exceedance at θ_j, with $\boldsymbol{\eta} = \{\eta_j\}_{j=1}^m = \boldsymbol{B}\boldsymbol{\beta}$ for B-spline basis matrix $\boldsymbol{B}$ and parameter vector $\boldsymbol{\beta}$ to be estimated. Roughness penalization of $\boldsymbol{\eta}$, with optimal roughness coefficient μ estimated by cross-validation, ensures good predictive performance. The penalized log-likelihood thus takes the form $\mathcal{L}_{pen}(\boldsymbol{\beta}) = \mathcal{L}(\boldsymbol{\beta}) - \mu \boldsymbol{P}$, with $\boldsymbol{P} = \boldsymbol{\beta}^\top \boldsymbol{D}^\top \boldsymbol{D} \boldsymbol{\beta}$ (cf., Section 5.3).

6.2.2 | Finite right endpoint

In case $\xi(\theta) < 0$, the limiting generalized Pareto distribution function has finite right endpoint. A consistent estimator for this right endpoint follows from (19) by setting $p = 0$
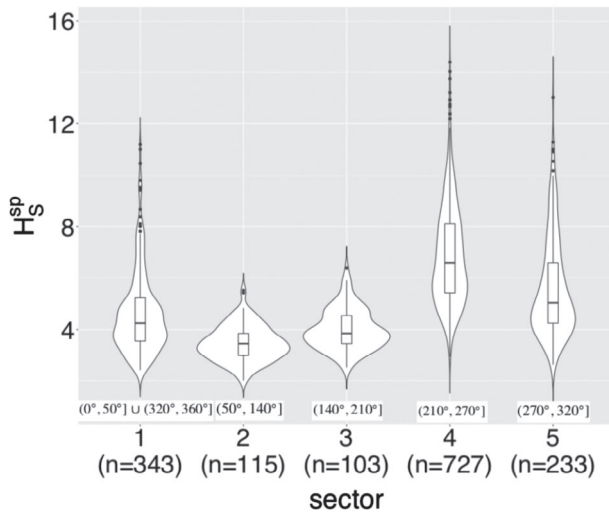
$$\hat{x}_0^{\mathrm{ML}}(\theta) = u(\theta) - \frac{\hat{\sigma}_u(\theta)}{\hat{\xi}(\theta)}. \tag{21}$$

7 | APPLICATION TO STORM PEAK SIGNIFICANT WAVE HEIGHT

The discussion above provides a suite of complementary approaches to nonstationary extreme value analysis incorporating elements of both parametric and semiparametric inference, featuring a common nonstationary extreme value threshold capturing covariate dependence of threshold exceedances. In this section, we apply the methodology to estimation of T-year values from the sample of storm peak significant wave height H_S^{sp} on storm direction θ introduced in Section 2. The mechanics of inference for extreme quantiles, and right endpoint if appropriate, is given in Algorithm 3 of Appendix B. For brevity henceforth, we refer to semiparametric, local parametric, and spline parametric estimators as "M", "local ML," and "spline ML," with "M" indicating a moment estimator, and "ML" referring to maximum likelihood.

Exploratory analysis of the sample, supported by previous analysis by Randell et al. (2016), suggests that the covariate domain can be partitioned into five directional sectors assumed approximately homogeneous in terms of the characteristics of H_S^{sp}. Referring to Figure 1, directional sectors corresponding to the following intervals of θ were identified. Sector 1 corresponds to $\theta \in (0°, 50°] \cup (320°, 360°]$, for storms propagating from the Norwegian Sea to the North; Sector 2 for $\theta \in (50°, 140°]$ corresponds to the "land shadow" of Norway, with fetch-limited storms propagating from the coast with a more northerly direction relative to the normal to the coast; Sector 3 is $\theta \in (140°, 210°]$, again for the Norwegian land-shadow, but with storms propagating from a more southerly direction; Sector 4 is $\theta \in (210°, 270°]$ corresponding to storms from the Atlantic potentially "funneled" by the Norwegian coast; and Sector 5 with $\theta \in (270°, 320°]$, for more northerly Atlantic storms. Further information about the underlying physics is given in Section 2. The partitioned sample is summarized in Figure 2, using so-called "violin" plots which add kernel density estimates to a box-whisker representation. The long-tailed behavior of storms from the Atlantic is clear in Sectors 4 and 5, compared with the fetch-limited characteristics in storms from Sectors 2 and 3. Although Sector 4 exhibits the largest values of threshold exceedances in Figure 2, there is evidence from the kernel density plots that Sector 5 has a relatively long tail. In this section we seek to quantify tail-heaviness by estimating $\xi(\theta)$ using both parametric and semiparametric approaches, and hence estimate extreme quantiles.

Using the approach in Section 4, nonstationary thresholds $u(\theta)$ were estimated using both the local ML and M estimator for lag $h = 60°$, concentration $\eta = 10$ and parameter $\phi = 0.35$; a sensitivity study was conducted to ensure that threshold inference was not overly sensitive to these choices. Estimates are shown together in Figure 3. The general trends shown by the two estimates are in agreement across the covariate domain; for example, larger values of $u(\theta)$ are observed for storms from the Atlantic (SW to W), and smaller values for fetch-limited directions (E to SE). Subsequent parametric and semiparametric inference for exceedance characteristics therefore has a relatively common starting point.

Estimates $\hat{\xi}(\theta_j)$ for $\theta_j \in \Theta$ from each of the M, local ML an spline ML approaches described in Section 5 are displayed in Figure 4, in terms of bootstrap means and 95% confidence intervals. Note that the M estimate of ξ was obtained using

FIGURE 3 Polar plot of extreme value thresholds $u(\theta)$ from the M (orange) and ML (blue) estimators. Storm direction θ is the direction from which waves propagate; radial scale of H_s^{sp} is in meters
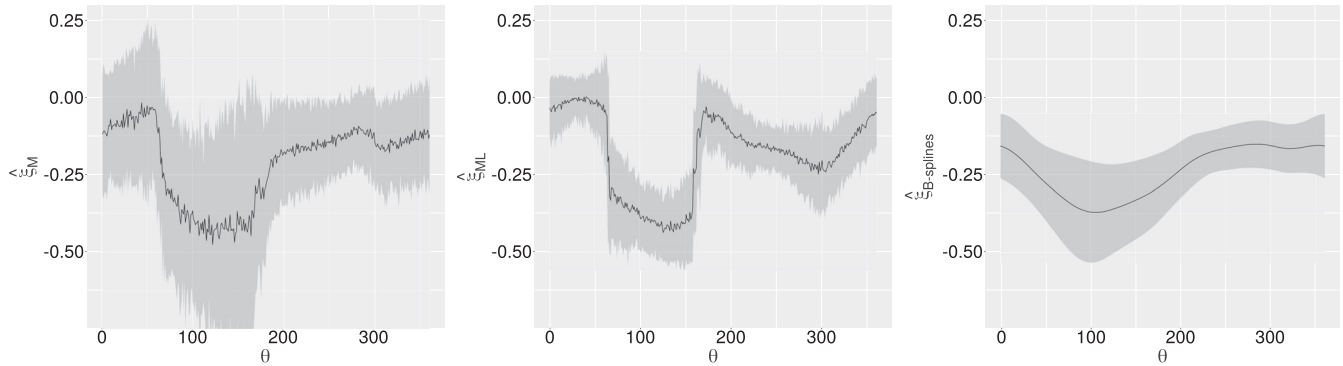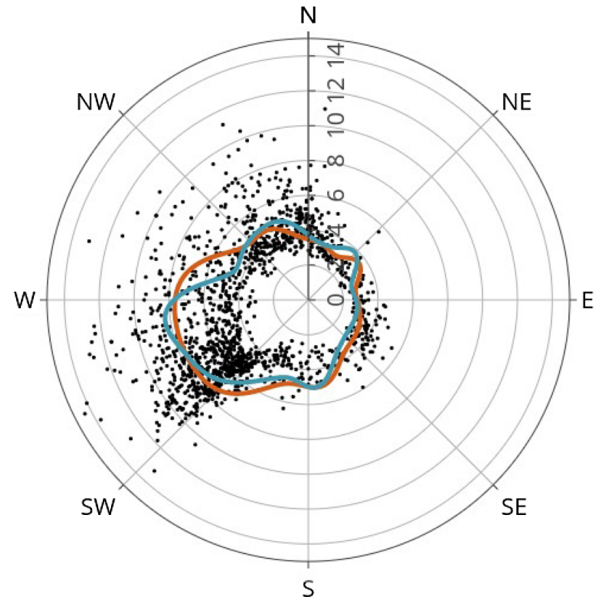

FIGURE 4 Bootstrap mean and corresponding 95% confidence intervals for $\hat{\xi}(\theta_j)$, $\theta_j \in \Theta$ based on: M (*left*), local ML (*middle*) and spline ML (*right*) estimators

the corresponding M estimate of u; analogously the local ML and spline ML estimates of ξ adopt the local ML estimate of u. There is reasonable overall agreement between the three estimates for ξ, and the estimates are also qualitatively plausible given other analyses of these data (Randell et al., 2016) and general physical considerations. Effects of land shadows (e.g., $\theta \in (80,150)$) resulting in lower values for ξ are clear. The spline ML estimator is smoothest with respect to θ and its confidence band does not include zero for any direction. In contrast, $\xi(\theta) = 0$ is suggested by semiparametric and local parametric estimators, most obviously for the M estimator, implying that the data generating distribution may lie in the Gumbel max-domain of attraction.

Identifying domains of attraction

Given the evidence in Figure 4, it is interesting to consider whether we can identify intervals with $\xi(\theta) < 0$, $\xi(\theta) = 0$ (and $\xi(\theta) > 0$ in principle) for $\theta \in S$ (cf., choice of domain of attraction, see Fraga Alves et al., 2017). We attempt this using two sets of hypothesis tests. In the first test, we adopt null hypothesis $F_\theta \in \mathcal{D}(G_0)$ versus the two-sided alternative $F \in \mathcal{D}(G_\xi)_{\xi \neq 0}$. Four different test statistics are considered, three of which are semiparametric

$$R_{k,n}(\theta) := \frac{\max\limits_{C(\theta,h)} \left(X_i(\theta_j) - u(\theta_j) \right)}{E^{(1)}(\theta)} - \log k(\theta)$$
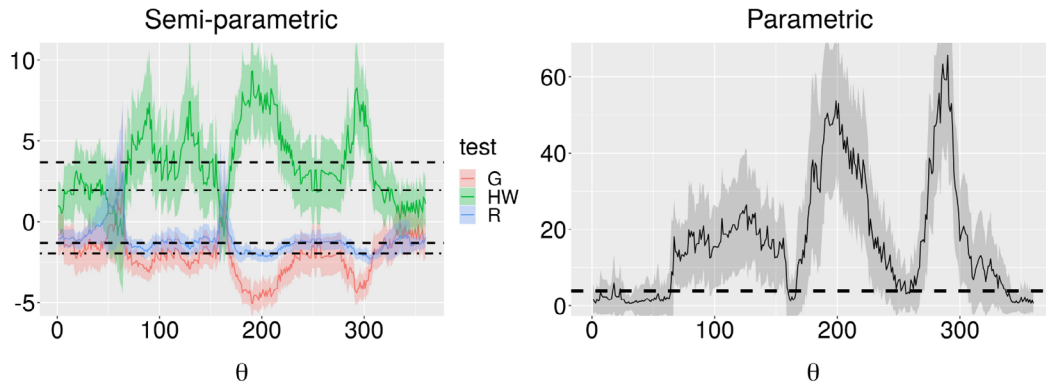
FIGURE 5 Bootstrap mean and corresponding 95% confidence intervals for the ratio (R, *R*), Greenwood-type (G, *G*), and Hasofer-Wang (HW, *W*) (left) and likelihood ratio (LR, *D*) (right) test statistics for $\xi = 0$. Horizontal lines in both panels indicate asymptotic critical values at significance level $\alpha = 5\%$. In the left panel: dashed lines for *R* test (-1.305 and 3.676); dot-dashed lines for *G* and *W* (± 1.96). In the right panel: dashed line at 3.84

$$G_{k,n}(\theta) := \sqrt{\frac{k(\theta)}{4}}\left(\frac{E^{(2)}(\theta)}{\left(E^{(1)}(\theta)\right)^2} - 2\right)$$

$$W_{k,n}(\theta) := \sqrt{\frac{k(\theta)}{4}}\left(\frac{1}{1 + \left[E^{(2)}(\theta)/\left(E^{(1)}(\theta)\right)^2 - 2\right]} - 1\right)$$

based on moment spacings

$$E^{(l)}(\theta) = \frac{1}{k(\theta)} \sum_{C(\theta,h)} \left(X_i(\theta_j) - u(\theta_j)\right)^l \mathbb{1}_{\{X_i(\theta_j) > u(\theta_j)\}} \quad \text{for } l = 1, 2$$

in the notation of Section 5; these test statistics are discussed in Fraga Alves et al. (2017) and references therein. The ratio test (R) assesses the relative contribution of the maximum to the average of the excesses over the given threshold *u*. For the R test, the null hypothesis $F_\theta \in \mathcal{D}(G_0)$ is rejected at level $\alpha \in (0, 1)$ if $R_{k,n}(\theta) < g_{\alpha/2}$ or $R_{k,n}(\theta) > g_{1-\alpha/2}$, where g_ε is the ε-quantile of the Gumbel distribution, $g_\varepsilon = -\log(-\log \varepsilon)$. The Greenwood-type test (G) and the Hasofer and Wang test (HW) are moment-based, and associated critical regions follow from the central limit theorem. The critical region for the two-sided test of nominal size α for $T^* \in \{G, W\}$ is given by $|T_{k,n}^*(\theta)| > z_{1-\alpha/2}$, with z_ε denoting the ε-quantile of the standard normal distribution.

When $\xi(\theta) > -1/2$, a parametric test statistic based on the generalized Pareto distribution is available (cf., Section 3.1 and Appendix A). Asymptotic normality of ML estimators for peaks over threshold gives rise to a likelihood ratio (LR) test using deviance statistic

$$D(\theta) = 2\left\{\mathcal{L}(\hat{\xi}(\theta), \hat{\sigma}_u(\theta)) - \mathcal{L}(0, \hat{\sigma}_u(\theta))\right\} \tag{22}$$

to test the hypothesis, where $\mathcal{L}$ is the log-likelihood in (13). At level α, the null hypothesis is rejected when the observed value for *D* exceeds the $(1 - \alpha)$-quantile of the χ_1^2 distribution (see, e.g., Coles, 2001).

Results using the four test statistics are illustrated in Figure 5. The values for test statistics G and HW lie generally on or beyond their critical values as a function of θ, providing evidence to reject the null hypothesis that $\xi(\theta) = 0$ at the significance level $\alpha = 5\%$. We note that upward crossings of the indicated $\alpha = 5\%$ critical barriers in the HW test sample path evidence an underlying short tailed distribution belonging the Weibull max-domain, whereas for the G-test, the Weibull max-domain is reflected in the downward crossings of corresponding $\alpha = 5\%$ critical barriers. The value of the R test statistic lies generally on or near its lower critical value, again providing some support for the Weibull max-domain of attraction; we note that this test statistic is recognized as being somewhat more conservative than G and HW, particularly for near-rejection in favor of the Weibull domain.

FIGURE 6 Bootstrap mean and corresponding 95% confidence intervals for the T-test statistic. The horizontal dashed lines correspond to the asymptotic critical values at the significance level $\alpha = 5\%$, that is, the normal quantiles which are solutions of $|z_\varepsilon| = 1.96$
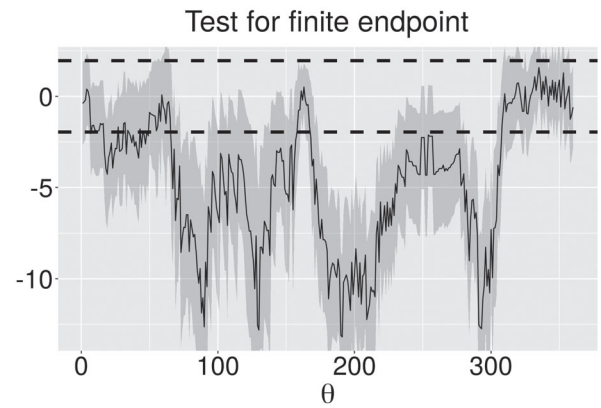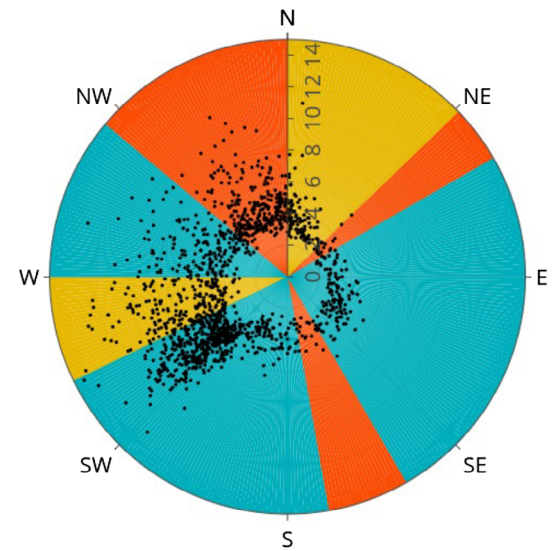

FIGURE 7 Polar plot showing combined deductions from tests of hypotheses for max-domain of attraction of H_s^{sp} at the $\alpha = 5\%$ significance level. Orange: $\xi(\theta) = 0, x^{F_\theta} = \infty$; yellow: $\xi(\theta) = 0, x^{F_\theta} < \infty$; blue: $\xi(\theta) < 0$. The original directional sample of H_s^{sp} on direction θ is also shown as black dots

The semiparametric approach provides a further possibility (Fraga Alves & Neves, 2014; Neves & Pereira, 2010) for hypothesis testing, specifically to test for a Gumbel max-domain of attraction with infinite right endpoint as null hypothesis, versus an alternative hypothesis of any max-domain of attraction with $\xi(\theta) \leq 0$ and $x^{F_\theta} < \infty$. No similar parametric test exists, since in the parametric setting a generalized Pareto distribution with $\xi(\theta) = 0$ corresponds to the exponential distribution which is right unbounded. In this sense, the semiparametric approach provides a better framework for tackling the statistical assessment of a finite upper bound. For every $\theta \in S$, we test H_0 : $\left(F_\theta \in \mathcal{D}(G_0) \text{ and } x^{F_\theta} = \infty\right)$ versus H_1 : $\left(F_\theta \in \mathcal{D}(G_\xi)_{\xi \leq 0} \text{ and } x^{F_\theta} < \infty\right)$ using the test statistic T given by

$$T_{k,n}(\theta) := \frac{1}{k(\theta)} \sum_{C(\theta,h)} \frac{X_i(\theta_j) - u(\theta_j) - \hat{a}_\theta\left(N/k(\theta)\right)}{\max_{C(\theta,h)}\left(X_i(\theta_j) - u(\theta_j)\right)} \mathbb{1}_{\{X_i(\theta_j) > u(\theta_j)\}}, \tag{23}$$

with M estimator $\hat{a}$ provided in (12). At the significance level $\alpha \in (0, 1)$, the null hypothesis is rejected whenever $\sqrt{k} \log k \, |T_{k,n}(\theta)| > z_{1-\alpha/2}$, where $z_\alpha = \Phi^\leftarrow(\alpha)$ denotes the α-quantile of the standard normal distribution function. Results using the test statistic T are shown in Figure 6. The null hypothesis of an infinite upper bound is rejected at the $\alpha = 5\%$ level for the majority of the directional domain, except notably for storms emanating from approximately $(310°, 360°)$, around $50°$ and around $160°$. From a physical perspective, these observations can be explained by occurrences of storms from the Norwegian Sea ($(310°, 360°)$), funneling from the north-east along the Norwegian coast ($50°$), and rare occurrences of dominant wave directions from the Norwegian coast ($160°$). We summarize the results of hypothesis tests concerning choice of max-domain of attraction in Figure 7, and note directional intervals of $\xi(\theta) = 0, \xi^{F_\theta} < \infty$ corresponding to storms from the Atlantic funneled along the Norwegian coast from south-west to west, and along the Norwegian coast from north to north-east.
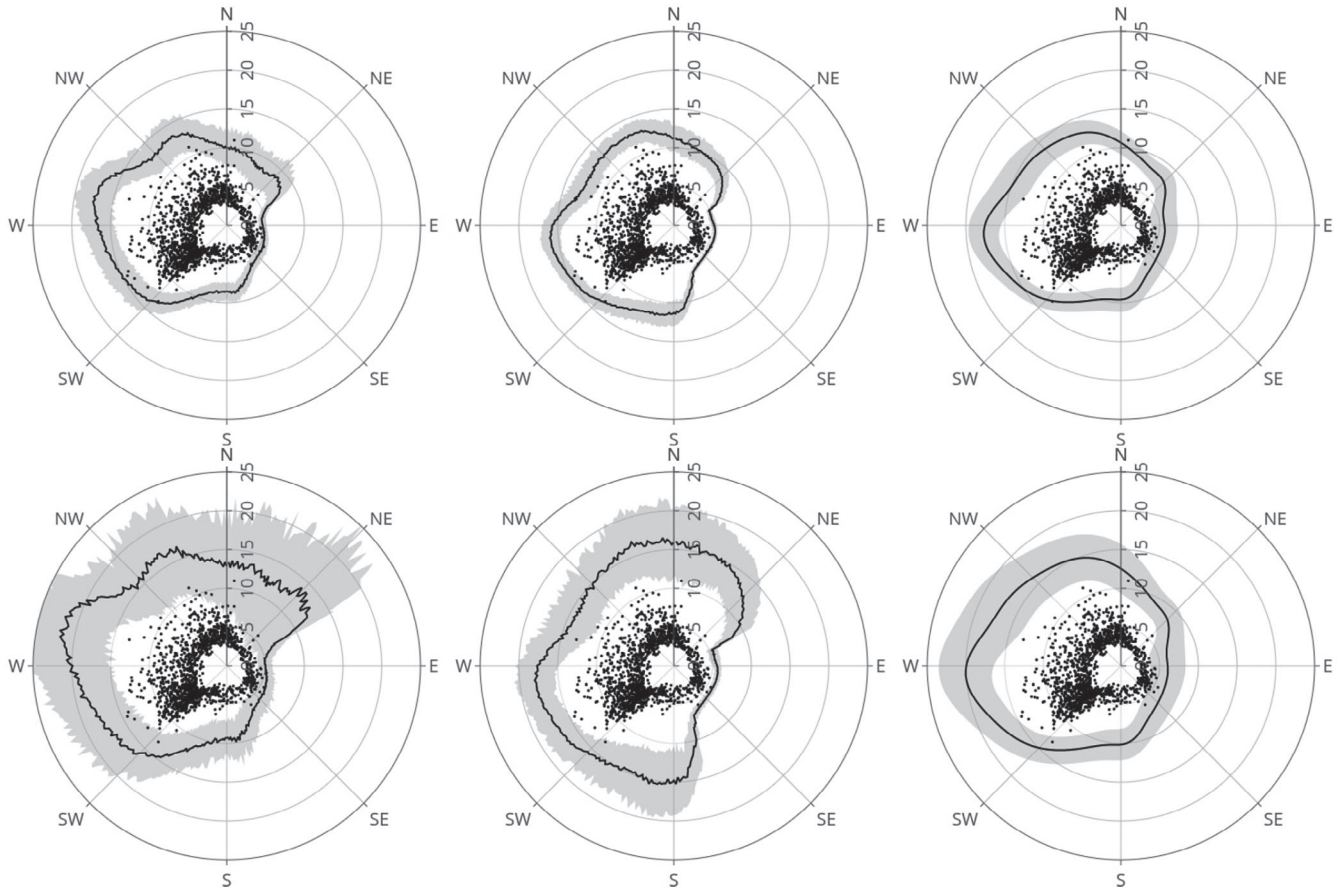
FIGURE 8 Rose diagrams for the 100-year (top) and 10,000-year (bottom) levels based on M (*left*), local ML (*middle*), and spline ML (*right*) estimators. Bootstrap means and corresponding 95% confidence bands are displayed as a function of direction measured clockwise from north

Estimating extreme quantiles and right endpoints

We use expressions (17) and (19) to estimate quantiles $x_p(\theta)$ using M, local ML and spline ML estimators. Assuming N_0 storm occurrences in observation period T_0, the T-year directional level corresponds to the value $x_p(\theta)$ such that $P\{X(\theta) > x_p(\theta)\} = (T_0/N_0) \times T^{-1}$. Figure 8 gives a matrix rose-plots of estimated directional 100-year and 10,000-year levels with accompanying 95% bootstrap confidence bands.

There is general qualitative agreement between the three estimates for 100-year level (top row) and 10,000-year level (bottom row), in terms of bootstrap mean. Uncertainties from the M estimation are somewhat larger, as might be expected recalling the evidence of ξ estimation in Figure 4. Not surprisingly, estimated extreme levels for directions with short fetches ($\theta \in [60,140)°$) are low, whereas those corresponding to long fetches from the Atlantic Ocean and Norwegian Sea ($\theta \in [310,360) \cup [0, 60)°$) are high. These findings are broadly in line with those from hypothesis testing in Figure 7. Estimated 100-year levels for H_s^{sp} are between 15 and 20 m the most severe sectors, in terms of bootstrap mean and confidence bands. The same is true of bootstrap means at the 10,000-year level except for M-related estimates which exceed 20 m. Randell et al. (2016) report 100-year values from a directional-seasonal parametric analysis of approximately 14 m for the Atlantic sector.

Figure 7 indicates there is significant evidence in the H_s^{sp} sample for concluding that the underlying distribution is right bounded on much of the directional domain. This suggests that estimation of the right endpoint x^{F_θ} might be usefully attempted. Three estimators are considered, as described in Section 6. The local ML and spline ML estimators for x^{F_θ} are only appropriate when $\xi(\theta) < 0$, whereas the semiparametric general estimator does not require estimation of $\xi(\theta)$ at all, but imposes the constraints $\xi(\theta) \leq 0$, $x^{F_\theta} < \infty$ by construction. To ensure fair comparison of the estimates for x^{F_θ}, we therefore repeated the local ML and spline ML generalized Pareto fits of Section 5, with the domain of ξ restricted
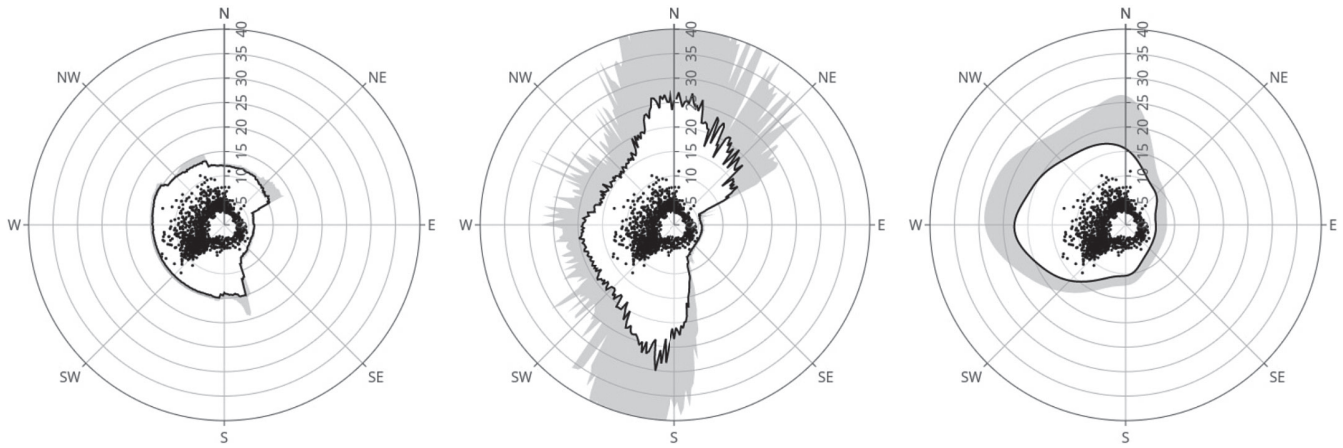
FIGURE 9 Rose diagrams for the finite right endpoint of H_s^{sp}. Three estimators are used: the general estimator (*left*), the local ML (*middle*) and spline ML (*right*) estimators. Bootstrap means and corresponding 95% confidence bands are displayed as a function of direction measured clockwise from north

to $\xi(\theta) < 0$. The left panel of Figure 9 shows estimates using general estimator (18). The center and right panels show estimates from local ML and spline ML estimation, with $\xi(\theta)$ constrained to be negative everywhere. Clearly, maximum likelihood estimation becomes more challenging as the true shape parameter ξ approaches zero from below, with numerical optimization routines more than often experiencing convergence issues (e.g., Gomes & Neves, 2008, for a comparison between M and ML within the univariate semiparametric setting). Estimates from the general estimator (left panel) are surprisingly low compared with those for local ML and spline ML, and with semiparametric estimates of the 100- and 10,000-year values using the moment estimator, shown in Figure 8. Agreement between the three estimators for the finite right endpoint cannot be expected to be nearly as good as that observed for T-year level, since the estimators are rather different in nature.

Local ML and spline ML estimates (Figure 9, center and right panels) show some agreement particularly for westerly, and north to north-easterly directions, where Figure 7 suggests $\xi(\theta) = 0$ and $x^{F_\theta} < \infty$; the local ML estimator provides considerably more uncertain estimates, because of input estimates for $\xi(\theta)$ near zero. The general endpoint estimator exhibits small variance in comparison, a distinctive trait that might be attributed to the simplicity in form not requiring external estimation of ξ; the small variance of this semiparametric estimator has been proved both theoretically and via extensive simulation studies (Fraga Alves & Neves, 2014). Increased uncertainty in the general endpoint estimator (Figure 9) corresponds to regions in Figure 7, for which $\xi(\theta) = 0$ and $x^{F_\theta} = \infty$. A study of the sensitivity of the general endpoint estimator to specification of h and $k(\theta)$ was undertaken, particularly due to the low estimates it yields. The study revealed that the general endpoint estimator is in fact remarkably stable with respect to both h and $k(\theta)$. We must of course keep in mind that all of the estimates in Figure 9 assume a finite right endpoint, whereas the evidence from Figure 7 is that this is not the case for all directions. Note further that the asymptotic distribution for the general endpoint estimator is nonnormal, and therefore the estimated bootstrap 95% confidence bounds offer a limited view of its asymptotic variance. A similar remark applies to the local-ML estimator for the right endpoint of the generalized Pareto distribution for the irregular case for $\xi \in (-1, -1/2]$.

In summary, application of the parametric and semiparametric methodologies developed in Sections 4–6 above to the sample of directional storm peak significant wave height suggests that estimates for ξ, and 100-year and 10,000-year levels are in good qualitative agreement. Where differences occur, they can be explained in terms of specific modeling assumptions made, rather than in terms of fundamental differences in the underlying approaches to extreme value analysis.

8 | DISCUSSION AND CONCLUSIONS

This article presents a framework for inference on nonstationary peaks over threshold, reconciling approaches from semiparametric and parametric extreme value analysis, in application to directional ocean storm severity. The key

components of the framework are (a) estimation of nonstationary extreme value threshold, and (b) estimation of tail characteristics from threshold exceedances, including extreme quantiles and finite right endpoint when appropriate. Threshold estimation is performed using a nonstationary extension of a heuristic approach for semiparametric moment (M) and parametric maximum likelihood (ML) estimators. Tail characteristics and extreme quantiles are then estimated based on semiparametric M, local parametric ML and spline ML estimators. Given threshold, the two streams of extreme value inference mirror each other. In the spline ML approach, a penalized B-splines representation with compact support is used: each basis function is nonzero on a specific interval of the covariate domain. This feature plays a similar role to bandwidth of directional neighborhoods in the semiparametric M and parametric local ML approaches.

We also consider estimators for the nonstationary finite right endpoint, including extension of the semiparametric general estimator of Fraga Alves and Neves (2014). Hypothesis tests, used to characterize the max-domain of attraction of storm peak significant wave height as a function of direction, suggest that the distributions of storm peak significant wave height for storms from the Atlantic and Norwegian Sea are unbounded above.

Inferences regarding directional thresholds for storm peak significant wave height are in good agreement over the covariate domain. Estimates for 100 and 10,000-year levels are also in reasonable agreement. Estimates for the right endpoint are more different across approaches, and are influenced by the specifics of modeling assumptions made associated with the different estimation strategies. For the application considered, both parametric and semiparametric inference provides similar characterizations of extreme nonstationary ocean environments. Indeed, we illustrate how ideas from the semiparametric and parametric schools of thought can be used in tandem to exploit the desirable features of the approaches, while overcoming some obvious pitfalls. For example, threshold estimation (used for both semiparametric and parametric analysis) is motivated by an inherently nonparametric heuristic in Section 4.

Parametric approaches to nonstationary extremes are relatively well studied due in part to the wide range of flexible covariate representations for generalized Pareto parameters for threshold exceedances, and associated methods for regression and assessment of model fit. In contrast, from a semiparametric perspective, an exact generalized Pareto correspondence need not be assumed, offering greater flexibility. However, avoiding a particular model choice from the outset generally results in increased uncertainty of estimates for ξ and high quantiles. Nevertheless, we show in this work that semiparametric and parametric approaches perform rather similarly when set up reasonably. We hope that the fusion of ideas from parametric and semiparametric approaches to extreme value analysis outlined in this article provide a basis for increased understanding and quantification of extreme phenomena in environmental and engineering applications.

ORCID
Evandro Konzen https://orcid.org/0000-0002-6275-1681
Cláudia Neves https://orcid.org/0000-0003-1201-5720
Philip Jonathan https://orcid.org/0000-0001-7651-9181

REFERENCES
Balkema, A. A., & de Haan, L. (1974). Residual life time at great age. *The Annals of Probability*, *2*(5), 792–804.
Chavez-Demoulin, V., & Davison, A. C. (2005). Generalized additive modelling of sample extremes. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, *54*(1), 207–222.
Coles, S. (2001). *An introduction to statistical modeling of extreme values*. Springer.
Coles, S., & Walshaw, D. (1994). Directional modelling of extreme wind speeds. *Applied Statistics*, *43*, 139–157.
Davison, A. C., & Smith, R. L. (1990). Models for exceedances over high thresholds. *Journal of the Royal Statistical Society: Series B (Methodological)*, *52*(3), 393–425.
de Haan, L. (1970). *On regular variation and its application to the weak convergence of sample extremes*. Mathematisch Centrum.
de Haan, L., & Ferreira, A. (2006). *Extreme value theory: An introduction*. Springer.
de Haan, L., Tank, A. K., & Neves, C. (2015). On tail trend detection: Modeling relative risk. *Extremes*, *18*(2), 141–178.
Dekkers, A. L. M., Einmahl, J. H. J., & de Haan, L. (1989). A moment estimator for the index of an extreme-value distribution. *The Annals of Statistics*, *17*(4), 1833–1855.

Ewans, K. C., & Jonathan, P. (2008). The effect of directionality on northern North Sea extreme wave design criteria. *Journal of Offshore Mechanics and Arctic Engineering*, *130*, 041604:1–041604:8.

Feld, G., Randell, D., & Jonathan, P. (2019). *On the estimation and application of directional design criteria*. Proceedings of the 38th International Conference on Ocean, Offshore & Arctic Engineering, ASME, Scotland. Retrieve from www.lancs.ac.uk/?jonathan.

Fisher, R. A., & Tippett, L. H. C. (1928). Limiting forms of the frequency distribution of the largest and smallest member of a sample. *Mathematical Proceedings of the Cambridge Philosophical Society*, *24*(2), 180–190.

Folgueras, P., Solari, S., & Losada, M. Á. (2019). The selection of directional sectors for the analysis of extreme wind speed. *Natural Hazards and Earth System Sciences*, *19*(1), 221–236.

Forristall, G. Z. (2004). On the use of directional wave criteria. *Journal of Waterway, Port, Coastal, and Ocean Engineering*, *130*(5), 272–275.

Fraga Alves, M. I., & Neves, C. (2014). Estimation of the finite right endpoint in the Gumbel domain. *Statistica Sinica*, *24*(4), 1811–1835.

Fraga Alves, M. I., Neves, C., & Rosário, P. (2017). Ageneral estimator for the right endpointwith an application to supercentenarian women's records. *Extremes*, *20*(1), 199–237.

Gnedenko, B. V. (1943). Sur la distribution limite du terme maximum d'une série aléatoire. *Annals of Mathematics*, *44*, 423–453.

Gomes, M. I., & Neves, C. (2008). Asymptotic comparison of the mixed moment and classical extreme value index estimators. *Statistics & Probability Letters*, *78*(6), 643–653.

Gumbel, E. J. (1958). *Statistics of extremes*. Columbia University Press.

Jonathan, P., & Ewans, K. C. (2007). The effect of directionality on extreme wave design criteria. *Ocean Engineering*, *34*, 1977–1994.

Jones, M., Randell, D., Ewans, K., & Jonathan, P. (2016). Statistics of extreme ocean environments: Non-stationary inference for directionality and other covariate effects. *Ocean Engineering*, *119*, 30–46.

Leadbetter, M. R. (1991). On a basis for peaks over threshold modeling. *Statistics & Probability Letters*, *12*(4), 357–362.

Mackay, E. B. L., Challenor, P. G., & Bahaj, A. S. (2010). On the use of discrete seasonal and directional models for the estimation of extreme wave conditions. *Ocean Engineering*, *37*, 425–442.

Mendez, F. J., Menendez, M., Luceno, A., Medina, R., & Graham, N. E. (2008). Seasonality and duration in extreme value distributions of significant wave height. *Ocean Engineering*, *35*, 131–138.

Neves, C., & Fraga Alves, M. I. (2004). Reiss and Thomas' automatic selection of the number of extremes. *Computational Statistics & Data Analysis*, *47*(4), 689–704.

Neves, C., & Pereira, A. (2010). Detecting finiteness in the right endpoint of light-tailed distributions. *Statistics & Probability Letters*, *80*, 437–444.

Northrop, P. J., & Jonathan, P. (2011). Threshold modelling of spatially dependent non-stationary extremes with application to hurricane-induced wave heights. *Environmetrics*, *22*(7), 799–809.

Northrop, P. J., Jonathan, P., & Randell, D. (2016). *Threshold modeling of nonstationary extremes*. In D. K. Dey & J. Yan (Eds.), *Extreme value modeling and risk analysis: Methods and applications* (pp. 87–108). New York: CRC Press, Taylor & Francis Group.

Oumow, B., de Carvalho, M., & Davison, A. C. (2012). Bayesian P-spline mixture modeling of extreme forest temperatures. Retrieved from www.mat.puc.cl/~mdecarvalho.

Pewsey, A., Neuhauser, M., & Ruxton, G. (2014). *Circular statistics in R*. Oxford University Press.

Pickands, J., III. (1975). Statistical inference using extreme order statistics. *The Annals of Statistics*, *3*(1), 119–131.

Randell, D., Turnbull, K., Evans, K., & Jonathan, P. (2016). Bayesian inference for non-stationary marginal extremes. *Environmetrics*, *27*, 439–450.

Reiss, R.-D., & Thomas, M. (2007). *Statistical analysis of extreme values* (3rd ed.). Springer.

Reistad, M., Breivik, O., Haakenstad, H., Aarnes, O. J., Furevik, B. R., & Bidlot, J.-R. (2011). A high-resolution hindcast of wind and waves for the North sea, the Norwegian sea, and the Barents sea. *Journal of Geophysical Research: Oceans*, *116*, 1–18.

Robinson, M. E., & Tawn, J. A. (1997). Statistics for extreme sea currents. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, *46*, 183–205.

Ross, E., Sam, S., Randell, D., Feld, G., & Jonathan, P. (2018). Estimating surge in extreme North Sea storms. *Ocean Engineering*, *154*, 430–444.

Smith, R. L. (1987, 09). Estimating tails of probability distributions. *The Annals of Statistics*, *15*(3), 1174–1207. https://doi.org/10.1214/aos/1176350499

Ugarte, M. D., Goicoa, T., Etxeberria, J., & Militino, A. F. (2012). Projections of cancer mortality risks using spatio-temporal P-spline models. *Statistical Methods in Medical Research*, *21*, 545–560.

Wood, S. N. (2017). *Generalized additive models: An introduction with R*. Chapman & Hall/CRC Press.

Zanini, E., Eastoe, E., Jones, M. J., Randell, D., & Jonathan, P. (2020). Flexible covariate representations for extremes. *Environmetrics*, *31*(5), e2624.

APPENDIX A. BASIS FOR INFERENCE ON DIRECTIONAL EXTREMES VIA THE POT METHOD

The contents of this appendix build on appendix B of de Haan et al. (2015). This article offers added flexibility to the latter by allowing $\xi = \xi(s)$ to vary with $s \geq 0$, where s might represent direction θ, or time or some other covariate. As a consequence, the right endpoint does not need to be assumed constant in s. We will not delve into the theoretical details in terms of explicit smoothness and boundedness conditions, needing to be in place particularly by assuming $h = h_n > 0$. These are clearly beyond the scope of this article, but we envisage that the probabilistic underpinning to this work will stem from chapter 9 of de Haan and Ferreira (2006).

For each direction $s \geq 0$, let $(X_1(s), X_2(s), \dots, X_n(s))$ be a vector of i.i.d. random variables with common distribution function $F_s(x)$, for all $x \in \mathbb{R}$, absolutely continuous with right endpoint $x^{F_s} \leq \infty$. Assume $F_s \in \mathcal{D}(G_{\xi(s)})$, for some $\xi(s) \in \mathbb{R}$ and for every $s \geq 0$, that is, that condition (1) holds locally for each s. In this setting, theorem 1.1.6 of de Haan and Ferreira (2006) ascertains that it is possible to replace n with t running over the real line in such a way that (1) becomes equivalent to the following extreme value condition: there exists a positive function a_s^* such that

$$\lim_{t \uparrow x^{F_s}} \frac{1 - F_s\left(t + xa_s^*(t)\right)}{1 - F_s(t)} = \frac{\log G_{\xi(s)}(x)}{\log G_{\xi(s)}(0)}, \tag{A1}$$

for all x with $1 + \xi(s)x > 0$. The limit in (A1) is the tail distribution function (also known as survival function) of the generalized Pareto (GP) distribution with shape parameter $\xi(s) \in \mathbb{R}$, given by $(1 + \xi(s)x)^{-1/\xi(s)}$. The extreme value condition (A1) is often key for describing rare events' behavior in lieu of the dual max-domain of attraction characterization $F \in \mathcal{D}(G_\xi)$.

(i) Parametric approach

Taking a parametric view, the peaks over threshold (POT)-domain of attraction condition (A1) prescribes the GP distribution as the proper fit to the normalized exceedances given these are above a certain high threshold near the right endpoint x^{F_s}. With minimal notational changes around a fixed (deterministic) threshold u, it is straightforward to see that condition (A1) implies

$$\lim_{u \uparrow x^{F_s}} \left| P\{X_1(s) \leq x + u \mid X_1(s) > u\} - H_{\xi(s),u,\sigma(u)}(x) \right| = 0, \tag{A2}$$

locally uniformly in $x > u$, for each $s \geq 0$, with $\sigma(u) > 0$ (we omit the subscript s on σ and μ below) for simplicity of notation) and $H_{\xi,\mu,\sigma}(x) := 1 - (1 + \xi(x - \mu)/\sigma)^{-1/\xi}$, for all x such that $0 < H_{\xi,\mu,\sigma}(x) < 1$, with location $\mu \in \mathbb{R}$ and scale $\sigma > 0$. Informally, $F_s^{[u]}(x) := P\{X_1(s) \leq x + u \mid X_1(s) > u\} \approx H_{\xi(s),u,\sigma(u)}(x)$, for all $x > 0$ and u near the right endpoint x^{F_s}, with the scale parameter implicitly defined in terms of s through the threshold $u(s)$.

For each s, the limiting relation (A2) provides the probabilistic underpinning for fitting a GP distribution function to the unconditional tail distribution function $\overline{F}_s(x) := 1 - F_s(x)$ with x sufficiently large. This becomes more evident since

$$F_s(x) = (1 - F_s(u))F_s^{[u]}(x) + F_s(u),$$

from which

$$\overline{F}_s(x) = \left(1 - F_s^{[u]}(x)\right)(1 - F_s(u)) \approx \left(1 - H_{\xi(s),u,\sigma(u)}(x)\right)(1 - F_s(u)), \tag{A3}$$

for $x > u$, as $u \to x^{F_s}$. Finally, we note that

$$\left(1 - H_{\xi(s),u,\sigma(u)}(x)\right)(1 - F_s(u)) = H_{\xi(s),\mu^*,\sigma^*(u)}(x), \tag{A4}$$

where $\mu^* - u = \sigma(u)\, U_H\left(1 - F_s(u)\right)$, $\sigma^*(u) = \sigma(u)(1 - F_s(u))^{\xi(s)}$, and U_H standing for the tail quantile function pertaining to the standard GPD, that is

$$U_H(t) := \left(\frac{1}{1 - H_{\xi(s),0,1}}\right)^{\leftarrow}(t) = \frac{t^{\xi(s)} - 1}{\xi(s)},$$

for all $t \geq 1$ (the left arrow indicates the left-continuous inverse). The representation (A4) facilitates the view that, in practice, changes in the threshold (e.g., through covariates) will be reflected in the scale parameter. In turn, the approach

for inference is reflected in the way we go choose to go about the term $1 - F_s(u)$ for this to become statistically meaningful. In order to able to perform large sample inference drawing on the POT-GPD framework streamlined above, we now make the threshold dependent on the sample size n and $u = u(n)$ will naturally become larger as n goes to infinity. A parametric approach typically advocates for a large enough threshold to be fixed and inference to be conducted on the basis of the resulting POT framework, whereby the expected number of exceedances above the selected threshold is a random number K_s, say, satisfying $(n/K_s)(1 - F_s(u_n)) \to 1$ in probability, as $n \to \infty$. This suggests estimation of $(1 - F_s(u))$ via the analogous tail empirical distribution function (stepping up by $1/n$ at each observation) evaluated at u, adding up to $1 - K_s/n$ in the above, associated with the random number K_s of exceedances of u at direction (or location) s. Hence, for a given (fixed) k_s, the location and scale parameters in (A4) become:

$$\mu^*(s) = u + \sigma(u) \ U_H\left(1 - \frac{k_s}{n}\right),$$

$$\sigma^*(u) = \sigma(u)\left(1 - \frac{k_s}{n}\right)^{\xi(s)}.$$

Therefore, the crux of parametric inference for extremes values lays in the estimation of the shape and scale parameters, respectively, $\xi(s)$ and $\sigma_s = \sigma^*(u)$.

(ii) Semiparametric approach
It will be notationally cleaner to express the argument (A1) in terms of the pertaining tail quantile function $U_s := (1/(1 - F_s))^{\leftarrow}$. Note that $U_s(t)$ is nondecreasing and provides a straightforward link to an extreme quantile, with the right endpoint representing the ultimate quantile: $\lim_{t \to \infty} U_s(t) = U_s(\infty) = x^{F_s}$. To this effect, we make t in (A1) depend on the (possibly unknown) sample size n at each location s through replacing it by $U_s(n/k_s)$, where k_s is an intermediate sequence of positive integers such that $k_s = k_s(n) \to \infty$ and $k_s/n \to 0$, as $n \to \infty$. This is possible because (A1) holds uniformly in x. Hence, we have for the left-hand side of (A1):

$$\frac{1 - F_s\left(U_s(n/k_s) + x\, a_s\left(U_s(n/k_s)\right)\right)}{1 - F_s\left(U_s(n/k_s)\right)} = \frac{n}{k_s}\left(1 - F_s\left(U_s(n/k_s) + x\, a_s^\star(n/k_s)\right)\right),$$

with $a_s^\star(n/k_s) = a_s\left(U_s(n/k_s)\right)$.

For simplicity, we consider regularly spaced independent vectors $(X_1(s), X_2(s), \dots, X_n(s))$, $s = 1, 2, \dots, m, \dots$ with i.i.d. components, with partial tally of $N = n \times m \in \mathbb{N}$ observations across the whole system, and where n is potentially unknown (without affecting inference on extremes), yet assumed large ($n \to \infty$). In this setting, the basic extreme value condition is (cf., (A3)):

$$\lim_{n \to \infty} \frac{N}{k\, \omega(s)} \left\{1 - F_s\left(U_s\left(\frac{N}{\omega(s)k}\right) + x\, a_s^\star\left(\frac{N}{\omega(s)k}\right)\right)\right\} = 1 - H_{\xi(s),0,1}(x), \tag{A5}$$

for all x with $1 + \xi(s)x > 0$, uniformly in $s = 1, 2, \dots$, with $\omega > 0$ a continuous and smooth function satisfying $(1/m)\sum_{s=1}^{m} \omega(s) \to 1$, as $m \to \infty$, that is, the resulting sequence of weights $\{\omega(m)\}_{m \in \mathbb{N}}$ is Cesàro summable. The latter is to maintain integrity, also ensuring that the stationary case is well defined. In particular, the case of complete stationarity, corresponding to omni-directional data in the context of this article, is recovered if $\omega(s)$ is identically one over the stipulated range for s. The interest lies in the estimation of the various extreme value indices $\xi(s)$, and the scale and location terms, respectively $a_s^\star(N/k_s)$ and $U_s(N/k_s)$, now with $k_s := [\omega(s) \times k]$ and $[\bullet]$ standing for integer part. Since the Nth-order statistic $X_{N-k_s:N}(s)$ is close to $U_s(N/k_s)$, if $k_s \to \infty$, $k_s/N \to 0$, as $N \to \infty$, we shall adopt it, as the usual estimator for the threshold $\hat{U}_s(N/k_s) = X_{N-k_s:N}(s)$. The random adaptive threshold in this setting emulates the nonstationarity mirrored in the scale $\sigma_s > 0$ which features the parametric setting (i).

The role of the of the (deterministic) function ω can perhaps be better understood if the domain of attraction condition (A5), characterizing nonstationary POT, is formulated in terms of their respective inverse functions intervening in both hand-sides. To this end, we use $t \to \infty$ in place of $N/k \to \infty$, so that (A5) translates into:

$$\lim_{t \to \infty} t \left\{1 - F_s\left(U_s\left(\frac{t}{\omega(s)}\right) + x\, a_s^\star\left(\frac{t}{\omega(s)}\right)\right)\right\} = \omega(s)\left(1 - H_{\xi(s),0,1}(x)\right),$$

which, by inversion in terms of $t/y = 1/\left\{1 - F_s\left(U_s(t/\omega(s)) + x\, a_s^\star(t/\omega(s))\right)\right\}$, and similarly for $y/\omega(s) = (1 + \xi(s))^{-1/\xi(s)}$ on the right-hand side, yields the desired condition:

$$\lim_{t \to \infty} \frac{U_s(t/y) - U_s(t/\omega(s))}{a_s^\star(t/\omega(s))} = \frac{(\omega(s)/y)^{\xi(s)} - 1}{\xi(s)},$$

for all $y > 0$. In particular, putting $y = 1$, and by virtue of $a_s^\star$ being a regularly varying function at infinity with index $\xi(s)$, uniformly in $s \in S$ (cf., de Haan & Ferreira, 2006), that is, $\lim a_s^\star(tx)/a_s^\star(t) = x^{\xi(s)}, x > 0$, we obtain the following condition on the "moving" threshold:

$$\lim_{t \to \infty} \frac{U_s(t/\omega(s)) - U_s(t)}{a_s^\star(t)} = \frac{(\omega(s))^{-\xi(s)} - 1}{\xi(s)}. \tag{A6}$$

In case $\xi(s) = 0$, for some s, then the limit in the above reads as $\log(\omega(s))$. Furthermore, if the right endpoint is finite, that is, $x^{F_s} < \infty$, then $a^\star > 0$ satisfies $a_s^\star(t) \to 0$, as $t \to \infty$, and $F_s \in \mathcal{D}(G_{\xi(s)})$, with $\xi(s) \leq 0$, which entails that $U_s(t)$ is asymptotic to $U_s(t/\omega(s))$, for $\omega(s) \in (0, \infty)$. The smooth function $\omega(s)$ primarily reflects the amplitude in extremes rather than their magnitude or severity.

APPENDIX B. ALGORITHMS

This appendix provides algorithms for estimation of local nonstationary threshold and optimal spline representation. It also provides an algorithm for the complete extreme value analysis given nonstationary extreme value threshold.

B.1 Local nonstationary threshold selection

Algorithm 1. Nonstationary threshold estimation for local estimators

---
1: Specify lag h, concentration η and parameter ϕ;
2: **for** $\theta_j \in \Theta$ **do**
3:     Estimate weights $\omega(\theta_{j'})$ for $\theta_{j'} \in \Theta$;
4:     **for** $k = 1, \ldots, N(\theta_j) - 1$ **do**
5:         Use the k largest values in $\mathcal{N}(\theta_j, h)$ to obtain $\hat{\xi}_k(\theta_j)$ using (8) or (6);
6:         Calculate $S_\phi(k)$;
7:     **end for**
8:     Set $k_j^* = \arg\min_k S_\phi(k)$;
9:     Identify threshold $u(\theta_j)$ with the $(k_j^* + 1)^{\text{th}}$ largest value with direction in $\mathcal{N}(\theta_j, h)$;
10: **end for**

B.2 Penalized maximum likelihood B-spline estimation

Algorithm 2. Maximum penalized likelihood for nonstationary periodic B-spline representation

---
1: Evaluate B-spline basis functions on index set Θ;
2: Specify sets of values of smoothing penalty coefficients λ and κ to consider;
3: **for** each choice of λ **do**
4:     **for** each choice of κ **do**
5:         **for** each of a number of bootstrap resamples **do**
6:             Estimate model parameters using bootstrap resample;
7:             Use estimated model to predict test observations (not occurring in bootstrap resample);
8:             Calculate the predictive log-likelihood;
9:         **end for**
10:         Accumulate total predictive log-likelihood;
11:     **end for**
12: **end for**
13: Select pair of values of λ and κ with best predictive performance, and evaluate spline coefficients for these choices of roughness coefficients.

B.3 Estimation of T-year level and finite right endpoint

The procedure for estimation of an extreme level and of the right endpoint, if appropriate, for any of the three methods is summarized in the following algorithm for clarity. The algorithm assumes that the nonstationary threshold $u(\theta)$, $\theta \in S$ has been predetermined through Algorithm 1 in connection with the adopted local estimators for ξ.

Algorithm 3. Extreme quantile (including T-year level) and finite right endpoint estimation using M, local ML, and spline ML based estimators

---

1: Specify data sample; non-stationary threshold $u(\theta)$; period of sample T_0, period T;
2: Isolate set of directional threshold exceedances;
3: **if** M **then**
4:     Specify window half-width h;
5: **else if** local ML **then**
6:     Specify window half-width h;
7: **else if** spline ML **then**
8:     Specify values of smoothing coefficients λ and κ to consider;
9:     Specify details for B-spline basis function construction;
10: **end if**
11: **for** each of a large number of bootstrap resamples **do**
12:     Generate bootstrap resample from sample of threshold exceedances;
13:     **if** M **then**
14:         Count number $k(\theta)$ of threshold exceedances in $\mathcal{N}(\theta, h)$;
15:         Count number $N(\theta)$ of observations in $\mathcal{N}(\theta, h)$;
16:         Estimate $\xi(\theta)$ on $\mathcal{N}(\theta, h)$ using (10);
17:         Estimate a_θ on $\mathcal{N}(\theta, h)$ using (12);
18:         Estimate high quantile with $p \ll 1/N(\theta)$ value using (17);
19:         Estimate finite right endpoint ($x^{F_\theta} < \infty$) using (18);
20:     **else if** local ML **then**
21:         Count number $k(\theta)$ of threshold exceedances in $\mathcal{N}(\theta, h)$;
22:         Count number $N(\theta)$ of observations in $\mathcal{N}(\theta, h)$;
23:         Estimate $\xi(\theta)$, $\sigma_u(\theta)$ on $\mathcal{N}(\theta, h)$ using (13);
24:         Estimate T-year value using (19);
25:         Estimate right endpoint (when $\xi(\theta) < 0$) using (21);
26:     **else if** spline ML **then**
27:         Estimate optimal smoothing parameters λ, κ and hence estimate $\xi(\theta)$, $\sigma_u(\theta)$ (Algorithm 2);
28:         Estimate optimal smoothing parameter μ and fraction $\tau(\theta)$ of threshold exceedances by logistic regression using (20);
29:         Estimate T-year value using (19), with $\hat{\tau}(\theta)$ in place of $k(\theta)/N(\theta)$;
30:         Estimate right endpoint (when $\xi(\theta) < 0$) using (21);
31:     **end if**
32:     Accumulate bootstrap estimates for parameters, extreme levels or quantiles and for the right endpoint;
33: **end for**
34: Calculate bootstrap means and confidence intervals for parameters, T-year values and endpoint.