Contour estimation using Penalised Piecewise Constant marginal and conditional extreme value models

Emma Ross[*], David Randell, Philip Jonathan

February 11, 2019

Abstract

This report describes the Penalised Piecewise Constant (PPC) model and software for estimation of environmental design contours using the conditional extremes model of Heffernan and Tawn [2004]. The sample is composed of peaks over threshold values for both a conditioning variate and its *associated* conditioned variates. Each pair is allocated to a particular *covariate bin*; all (joint) observations with the same covariate bin are assumed to have common extreme value characteristics. The non-stationary marginal extreme value characteristics of each variate is estimated using roughness-penalised maximum likelihood estimation using a generalised Pareto (GP) model above the threshold and gamma below. The extremal dependence structure between the variates on a transformed standard scale (Gumbel or Laplace) is then estimated using a conditional extremes model, also piecewise non-stationary with respect to covariates. Different approaches to contour estimation, generally reliant on simulation under the fitted models, are outlined.

Major Updates Since Previous Release

- Extended to more than one covariate and non-periodic covariates;
- Hefferenan and Tawn model extended from bivariate to multivariate cases;
- Marginal model now fits a Gamma distribution below the threshold, instead of an empirical distribution;
- Margins can now be transformed to either Laplace or Gumbel scale (the Laplace scale more naturally handles negative dependence);
- Improvements to memory usage and speed of return value estimation;
- Use of importance sampling to speed up computation of contours;
- Alteration to contouring options: empirical and Heffernan and Tawn (H&T) density contouring methods merged; radial quantile method replaced with Huseby contour (similar concept but more rigorously defined);
- Contour levels are derived from the same return periods used in the Marginal and H&T stages, i.e. no longer defined by probability level on the conditioned variable

*Corresponding author. Email: `e.ross@shell.com`

Contents

0 Introduction

The conditional extremes model of Heffernan and Tawn [2004], and extensions such as Jonathan et al. [2014], Keef et al. [2013] provide a framework to estimate multivariate extremal dependence in the presence of covariates, and hence to estimate design contours and other statistics of interest in metocean design. The approach is motivated by an asymptotic form for the limiting conditional distribution of one or more conditioned random variables given a large value of a conditioning variable. An outline of the approach is given by Jonathan et al. [2010]. Conditions for the asymptotic argument to hold have been explored by Heffernan and Resnick [2007].

Suppose we want to estimate design contours using the conditional extremes model for bivariate peaks over threshold of random variables $\dot{Y}_1$ and $\dot{Y}_2$. These variables might be significant wave height H_S and associated peak period T_P, and the dependence between them might be non-stationary with respect to covariates $\boldsymbol{X}$ such as season or storm direction. We then want to simulate realisations under the model, and use the simulation to estimate design contours. We propagate uncertainties due to tuning parameter choice (specifically, threshold levels for marginal and dependence models) and sampling throughout the inference, so that design contours reflect these.

Non-stationarity with respect to covariates $\boldsymbol{X}$ is captured in the model using a Penalized Piecewise Constant (PPC) approach. Namely, covariates are split into bins considered to be roughly homogeneous. Model parameters are then estimated as constants within each covariate bin. To avoid over-fitting, a penalty on the parameter difference between covariate bins can be imposed, e.g. for the non-stationary GP scale parameter. An appropriate value for this roughness penalty is estimated using k-fold cross validation. Other parameters, such as the rate of concurrence and threshold, are simply estimated independently per covariate bin.

Below we give an overview of the 5 stages of analysis included in the PPC software. Note that these stages are described here in terms of the simplest application of the software, namely using one conditioned and one associated response varying with respect to a single covariate. The approach extends easily to a higher number of associated responses and covariates, and indeed the software can be used for such analysis.

1. The first stage deals with finding or simulating storm peaks. Peaks are chosen using the main (conditioning) variable $\dot{Y}_1$. $\dot{Y}_2$ is then the associated value at the time of the storm peak in $\dot{Y}_1$. This is discussed in section 1.

2. Secondly the user splits the data into bins based on the marginal response characteristics. This is discussed in section 2.

3. The marginal PPC models for the distribution of $\dot{Y}_1$ and $\dot{Y}_2$ are then fitted in turn and are used to transform response data on the original scale, $\{\dot{y}_{1i}, \dot{y}_{2i}\}_{i=1}^{N}$, to data on Gumbel or Laplace scale, $\{y_{1i}, y_{2i}\}_{i=1}^{N}$. Details of this step are discussed in section 3.

4. We then fit a conditional extremes model for $Y_2|Y_1$ for various choices of threshold for the conditioning variate, retaining the estimated model parameters and residuals. This is discussed in section 4.

5. Finally, we estimate design contours. Using the output from the marginal and conditional extremes model, Monte Carlo simulations are run and contours drawn using various methods. Details of this step are discussed in Section 5.

   This software was developed as part of a project part-funded by the European Union ERANET entitled Environmental Contours for SAfe DEsign of Ships and other marine structures" (EC-SADES), along with a review paper on the definition and application of environmental contours [Ross et al., 2019]. Further, this software is applied to analysis of surge in the Northern North Sea and described in detail in Ross et al. [2018].

1 Stage 1: Data preparation

In Stage 1 we prepare D-dimensional peaks over threshold response data $\{y_{1i}, y_{2i}, \dots y_{Di}\}_{i=1}^{N}$ with C associated covariates $\{x_{1i}, x_{2i}, \dots, x_{Ci}\}_{i=1}^{N}$. The user has a choice of 2 different run files for this stage:

- `Stage1_PeakPicking.m`
- `Stage1_SimulateData.m`

When data is available, for example time series observations for H_S and T_P, `Stage1_PeakPicking` should be run to find storm peaks from the full time-series. In the case that we have no existing data, or we simply want to test the model, `Stage1_SimulateData.m` should be run to generate response data directly.
Output data: `Output\Data.mat`

Figures generated when running the software are stored in a 'Figures' subdirectory, with a prefix `{StgX}` linking the figure to the Stage it came from.

1.1 Stage1_PeakPicking

This script converts time series data into peaks-over-threshold data, suitable for modelling with the generalised Pareto distribution. The name of the Matlab data file should be provided to the load() command. Notes are provided below for the key inputs to this script.

- `Rsplbl` [$D\times1$] cell array, containing string descriptions/names for the main and associated responses (in that order) - ensures that plots produced by the analysis are labelled correctly.
- `CvrLbl` [$C \times 1$] cell array containing string descriptions/names for the covariate(s).
- `Rsp` [$N \times 1$] vector containing the main response data (the response which we condition on).
- `Cvr` [$N \times C$] matrix where each column contains different covariate data. Number of columns must match size of `CvrLbl`.
- `IsPrdCvr` [$C \times 1$ boolean] flag dictating periodicity of covariate(s). If 1, covariate data loops on 360. When using more than one covariate, this is a vector input with one flag per covariate, e.g. [1,0]. Note that, if you have a periodic covariate which is *not* on [0,360), you must rescale it to cover this range to enable periodicity to be accounted for. Non-periodic covariate data can, on the other hand, be provided on any scale.
- `Asc` [$N \times (D-1)$] matrix where each column contains a different associated response - the responses which will be conditioned on the value of the main response given in `Rsp`. Number of columns must match the size of `RspLbl` minus one.
- `NEP` [scalar] non-exceedance probability (on [0,1)) used to define the threshold for storm-identification.

Suppose that we have set the main response `Rsp`= H_S and associated response `Asc`= T_P; the identification of storm trajectories and storm peak exceedances is illustrated in Figure 1. Note that we peak pick over the main response `Rsp` (in this case H_S) and take *associated* observations as peaks over threshold for T_P. The Matlab data output of this script (as well as the alternative `Stage1_SimulatedData.m`) is saved as `Output\Data.mat` which has the following format:

- `Dat.Y` [$N \times D$] matrix containing response data, with the main response (the one we condition on) in the 1st column, and associated in the subsequent columns;
- `Dat.X` [$N \times C$] matrix containing covariate data
- `Dat.RspLbl`: [$1 \times D$] cell array containing string descriptions of the data stored in the columns of `Dat.Y`;
- `Dat.CvrLbl`: [$1 \times C$] cell array containing string descriptions of the data stored in the columns of `Dat.Y`;

- `Dat.IsPrd`: $[1 \times C]$ vector of 0s and 1s dictating periodicity of the covariates in `Dat.X`.

An example of storm peak-picked data is shown for North Sea data in Figures 2 and 3. A quantile level of $\tau = 0.6$ was used to set the peak picking threshold giving 2566 storms.

1.2 Stage1_SimulateData

An alternative to using real data is to test the model using simulated data with known characteristics. Note that, though this update to the code accommodates *fitting* models for multivariate cases with multiple covariates; the simulation script is restricted to bivariate cases with a single covariate only. The first four inputs required by the user set the dimensions of the data to be simulated:

1. `nDmn`: the number of response variables you want to simulate
2. `nObs`: the number of observations you want to simulate
3. `nBin`: the number of covariate bins you want (common to both margins if `nDmn` > 1)
4. `BinEdg`: vector of edges of covariate bins in [0,360] (these will wrap around 0)

For each response, the user is then required to set the following distributional properties based on the number of bins `nBins` you specified:

1. `MM.Shp` : GP shape parameter
2. `MM.Scl` : GP scale parameter (vector of length `nBin`, a shape parameter for each covariate bin)
3. `MM.Thr` : GP threshold (vector of length `nBin`, a threshold for each covariate bin)
4. `Rat` : Poisson rate (again a vector of length `nBin`)

Finally, in the case that the user chooses to simulate two responses (`nDmn` $= 2$), the dependence model used and its associated parameters should also be set with the following inputs:

1. `Jnt.Mth`: Choice of dependence model: multivariate normal `MVN`; logistic `LGS`; or asymmetric logistic `ASL`
2. Associated parameters:
   - `MVN` : dependence parameter `Rho` $\in [0, 1]$
   - `LGS` : dependence parameter `Alp` $\in [0, 1]$
   - `ASL` : dependence parameter `Alp` as above and weighting parameters `Theta` (one for each response/margin) $\in [0, 1]$ setting the proportion of 'random' points off of the logistic dependence

The result of running this script is the `Output\Data.mat` file as described in the previous section. Figure 4 provides an example, akin to the black-dot peak observations in Figure **??**.

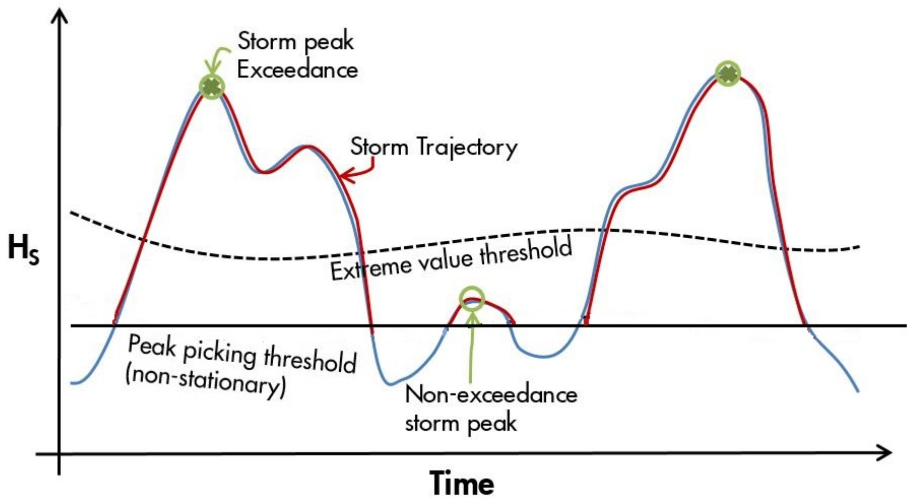

Figure 1: Peak Picking Illustration

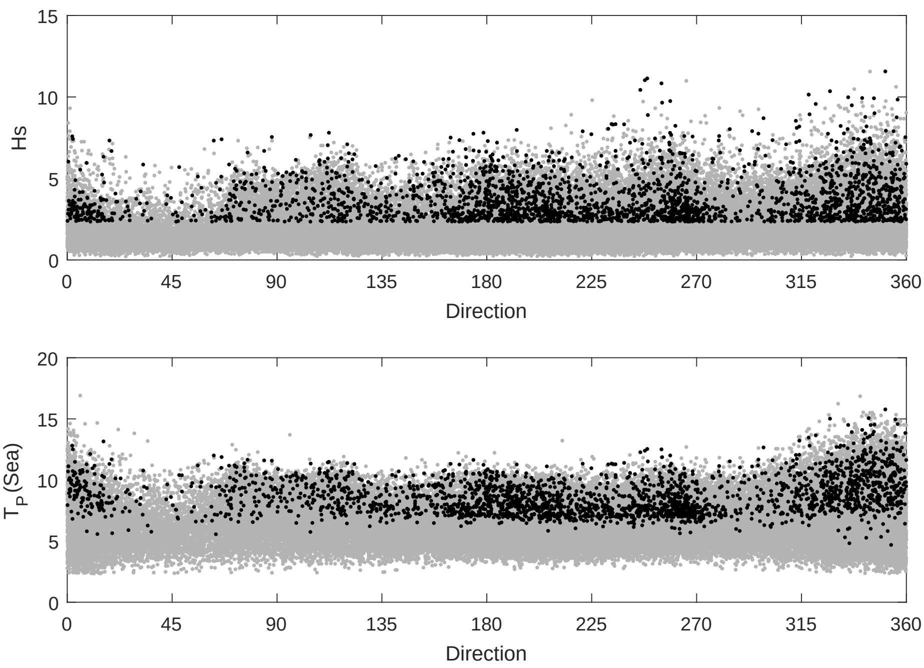

Figure 2: Marginal Hs and Tp as a function of direction for North Sea data. Storm peaks shown in black, all sea states in grey

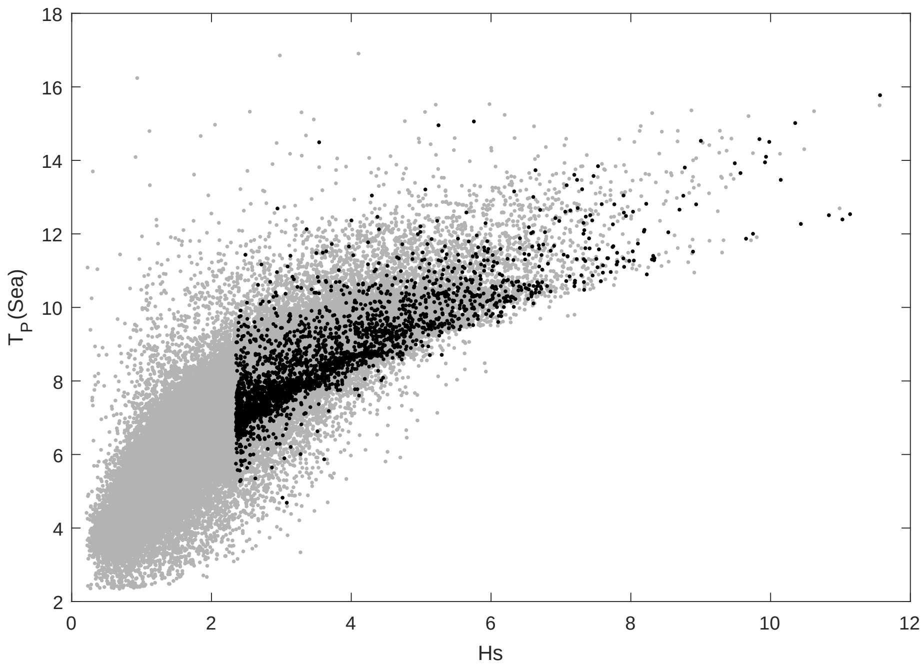

Figure 3: Joint distribution Hs and Tp for North Sea data. Storm peaks shown in black, all sea states in grey

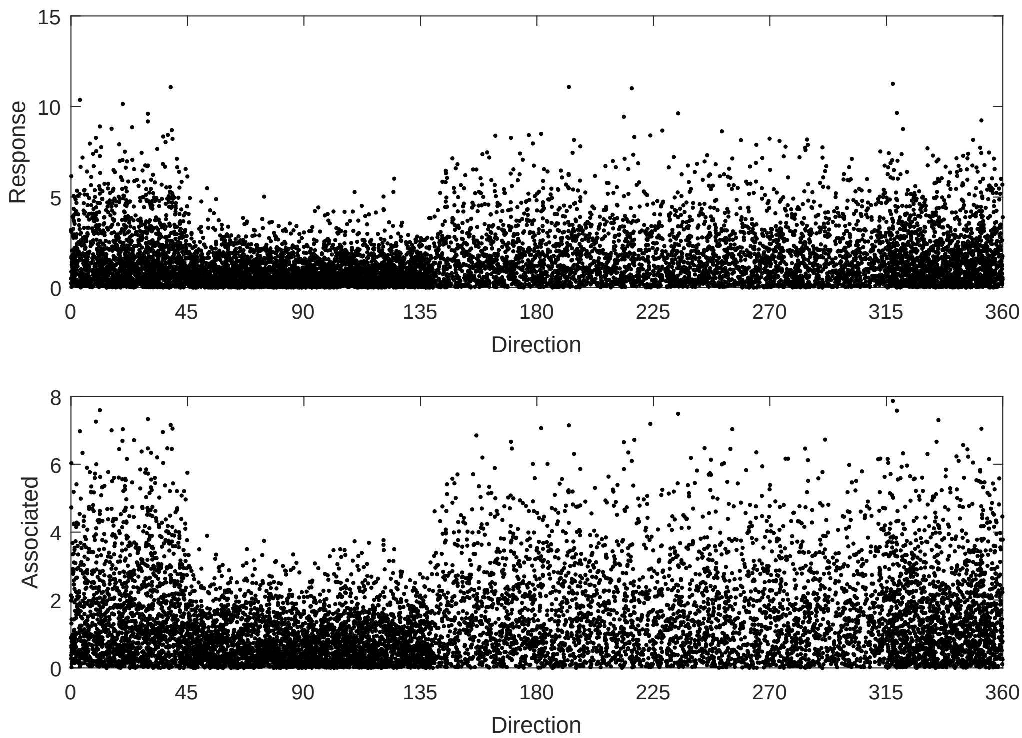

Figure 4: Example Simulated Data

2 Stage 2: Choose covariate bins

At the start of Stage 2, peak-picked data $\{y_{di}\}_{i=1}^{N}$ for each marginal response $d \in \{1, \ldots, D\}$ are loaded from Stage 1. In order to fit a piecewise constant Gamma-GP model to this data, we first need to specify covariate bins `BinEdg`. This script is used to experiment with and set covariate bin edges. A plot of the marginal data against the covariate(s) with current bin locations marked in red is produced. The goal is to set bin edges which effectively separate the data into sections with homogeneous covariate characteristics (rate and scale), after which we move onto Stage 3. As soon as you are happy with your bin choice, you can move onto Stage 3. The set of bin-edges you last tried will automatically be fed to Stage 3 for subsequent use.

2.1 Running Stage2

Run Script: `Stage2_SetBinEdges.m`
Output files: `Output\Bin.mat`

The only user-input to this stage is the $1 \times C$ cell array, `BinEdg`, storing bin edges for each covariate. In the case of a single covariate, bin edges should be provided in $\{[...]'\}$ format. Note that we need to transpoe ($'$ operation in Matlab) to put the bin edges into long vector format. In the case of multiple covariates; bin edges should be provided in $\{[...]', [...]'\}$ format; resulting in 2D bins which are the multiplicative combination of bins in each individual covariate dimension.
Note that, for covariates identified as periodic (setting `IsPrdCvr` $= 1$ in Stage1), bins will automatically wrap around 360. This means that, if 0 or 360 are not specified in the vector of bin edges for that covariate; by default there will be a bin which straddles 0. If your covariate data is periodic but not on $[0, 360)$, you will need transform it to $[0, 360)$, e.g. by adjusting the raw data within the Stage2 script, before assigning it to `Cvr`.
If you are using a non-periodic covariate, the data can be on any scale but bear in mind that the first and last entries in the bin-edge vector will be interpreted as end-points for the range of the covariate. Specifically, you should take care to ensure that the outer bin edges (first and last) are wider than the range of the data. If you do not do this, an error will be produced when a check is made that the data lies within the range defined by the first and last bin edges.
Further points to note:

- A warning will be produced if you have too few observations (< 30 total number of observations, not exceedances) in any given bin. This is to ensure you have enough data to fit to in each bin and prevents you from over-fitting by defining too many bins.
- If the total number of bins in the model is > 16, the code can struggle to estimate the generalised Pareto model well, so the number of bins should be kept relatively small. Note that a small number of bins in each covariate dimension multiplies to a large number of total bins; e.g. 4 bins in direction and season results in $4 \times 4 = 16$ bins in total.
- This code is designed to run non-stationary models and hence expects some form of covariate input. A non-stationary (covariate-free) model can be run using this code however, by creating a single periodic bin via:
  - supplying e.g. time or an index on the observations to the `Cvr`;
  - setting `IsPrd` to 0 for all covariates (enforcing periodicity);
  - setting `BinEdg` to $\{[\min(\texttt{Cvr}), \max(\texttt{Cvr})]'\}$.

  The user is however **strongly encouraged** to incorporate covariates which are known to strongly affect the response(s). Failure to account for covariate effects can lead to very different return-value estimates and environmental contours.

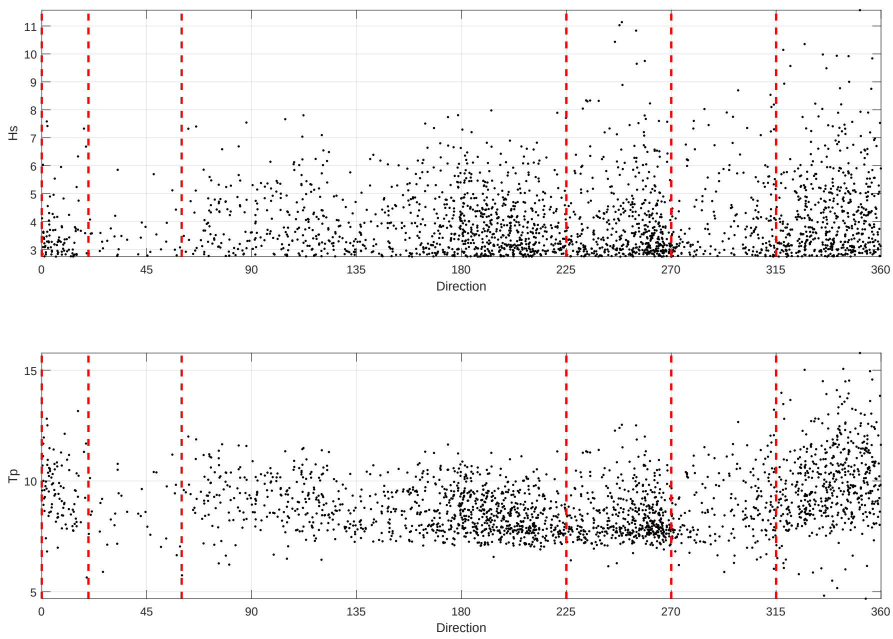

Figure 5: Example bin allocation. Bins are chosen at [0,25,60,230,275,315] degrees. Storm peak data shown in black, chosen bin edges are shown with red dashed lines.

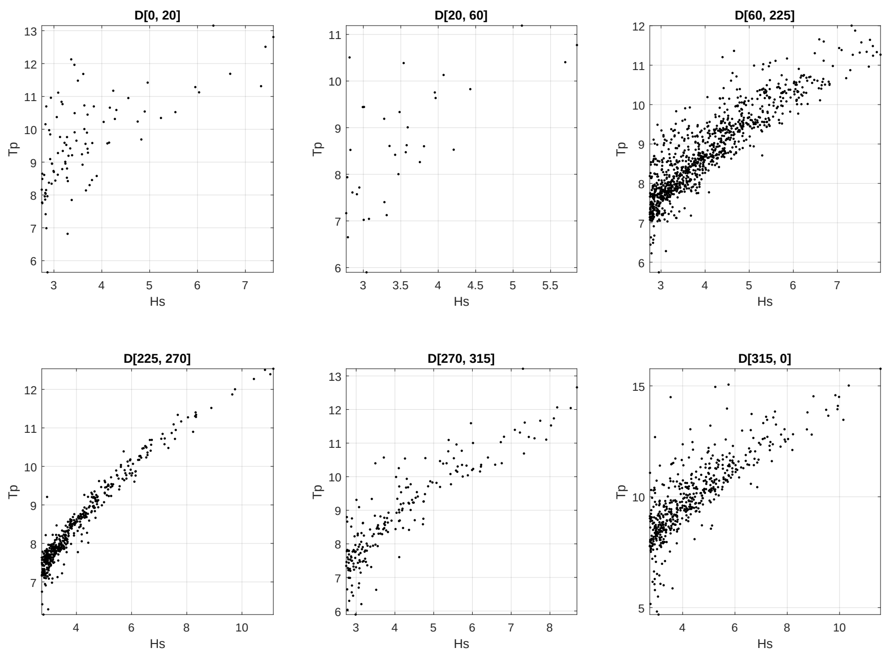

Figure 6: Scatter plots of storm peak responses broken out by bin.

3 Stage 3: Fit marginal PPC models

3.1 The penalised piecewise constant model

Stage3 fits a PPC extreme value model to a single marginal response. Non-stationary marginal extreme value characteristics of each variate are estimated in turn using a Gamma-GP model (GP above the threshold, Gamma below) and roughness-penalised maximum likelihood estimation. For a given variable and covariate bin k, the extreme value threshold $\psi_k(\tau)$ is assumed to be a quantile of the Gamma distribution fitted to all data in that bin, with specified non-exceedance probability τ. τ is constant across bins. Threshold $\psi_k(\tau)$ is estimated with no smoothing across bins.
Threshold exceedances are assumed to follow the GP distribution with shape $\xi(\tau)$ and scale $\nu_k(\tau)$. Since estimation of the shape parameter is particularly problematic, the shape parameter is assumed constant (but unknown) across covariate bins. The extent to which the GP scale varies across bins is controlled by smoothness parameter λ. Then parameters $\xi(\tau), \{\nu_k(\tau)\}$ are estimated using penalised (log-) likelihood optimisation, maximising the value of the likelihood given in Appendix A.
Data below the threshold is assumed to follow a 3-parameter Gamma distribution with parameters location $l_k(\tau)$, shape $\omega_k(\tau)$, and scale $\kappa_k(\tau)$, all piecewise constant with respect to covariate bins. The density and cumulative distribution function for this non-standard parametrisation of the Gamma distribution is provided in Appendix A.2. Note that the extent to which the Gamma parameters vary by bin is *not* controlled by any smoothness parameter (whereas the GP scale's smoothness is controlled).
The Poisson rate of storm occurrence, GP scale and threshold vary across bins but are constant within each bin.
For the given margin, the resulting PPC model is used to transform the response data on original scale to standard margins (Gumbel or Laplace, user-chosen in the Stage3 script) using the Probability Integral Transform (PIT) per covariate bin. Details of this procedure can be found in the Appendix. Laplace scale is generally preferred as it permits negative conditional dependence. In the presence of negative dependence, if the user wants to use the Gumbel distribution they must first flip the sign of one of the variables to define a positive-dependence problem.

3.2 Uncertainty quantification

Two sources of randomness are carried through the estimation procedure. Firstly, the model is fitted for multiple bootstrap samples of the data, uncertainty in the resulting model parameters then being carried through to later modelling stages. Secondly, for each bootstrap sample, the marginal non-exceedance probability (used to establish the threshold within each covariate bin) is randomly sampled from a range provided by the user.

3.3 Running Stage3

Run Scripts: `Stage3_FitMargin1.m`, `Stage3_FitMargin2.m`
Output files: `Output\MM1.mat`, `Output\MM2.mat`

Inputs
The following inputs are listed in order of usage. Note that the parameters which you will most likely need to tune/play with are `NEP`, `CV.SmthLB` and `CV.SmthLB`.

- `iDmn` [scalar] specify the response upon which to fit marginal model
- `NEP` [1×2] non-exceedance probability range, should be in $(0, 1)$
- `nB` [scalar] number of bootstrap re-samples - must use same number for each margin
- `Yrs` [scalar] number of years the data spans
- `RtrPrd` [$1 \times R$] vector of return periods (years)
- `CV.CVMth` [boolean] If 0: only Cross Validate smoothness parameter for original dataset (fast); or 1: Cross Validate smoothness for every bootstrap resample (slow)

- CV.nCV [scalar] number of cross-validation groups
- CV.nSmth [scalar] number of smoothness parameter values tried in cross-validation
- CV.SmthLB [scalar] lower bound (log10) for smoothness range
- CV.SmthUB [scalar] upper bound (log10) for smoothness range
- MarginType [string] specify the standard margin scale on which the Heffernan & Tawn model will be fitted (options are 'Laplace' or 'Gumbel')

This stage should be run at least twice; specifically, once for each margin. To keep track of the input settings used for each margin and to ensure you've fitted a marginal model for each response, it is good practice to keep nDmn (=D) copies of the Stage3_FitMargin.m script (e.g. as we have listed under 'Run Scripts:' above). If you forget to fit a marginal model to one of your responses, you'll typically face the following error when running Stage 4: `Meg should be an nDmn x 1 Marginal Model`.
Note that the input settings for each margin can generally differ, however the number of bootstraps nB and standard margin MarginType must be common to all scripts, for consistency when fitting the conditional model in Stage 4.
Outputs
Each marginal model crated a mat file MMi for the *ith* dimension run. There should be one for each marginal once Stage 3 has been run

- MM.X [$N \times C$] the covariate data
- MM.Y[$N \times 1$] the observational data
- MM.Yrs [scalar] the number of years of data
- MM.RspLbl [string] the label for the response modelled (used in plots)
- MM.RspSavLbl [string] the label for the response modelled (used in saving files)
- MM.CvrLbl [string] the covariate labels
- MM.nB [scalar] number of bootstraps used
- MM.RtrPrd [$1 \times R$] return periods
- MM.Bn] covariate bin structure (created in Stage 2)
- MM.Scl [$K \times nB$] Generalised Pareto Scale parameter
- MM.Shp [$nB \times 1$] Generalised Pareto shape parameter
- MM.Omg [$K \times nB$] Gamma parameter
- MM.Kpp [$K \times nB$] Gamma parameter
- MM.GmmLct [$K \times 1$] Gamma location parameter
- MM.NEP [$K \times nB$] non exceedance probability
- MM.Thr [$K \times nB$] exceedance threshold
- MM.Rat] [$K \times nB$] Rate of occurrence
- MM.BSInd [$N \times nB$] index vector for bootstrap reordering
- MM.nCvr [scalar] number of covariates in the model
- MM.nDat [scalar] number of observations
- MM.nRtr [scalar] number of return values R

- `MM.RVPrb` $[(K+1) \times nRVX \times R]$ return value probabilities CDF the final bin is the Omni return value CDF
- `MM.RVX` $[nRVX \times 1]$ location at which return value CDF has been computed
- `MM.RVMed` $[(K+1) \times R]$ median return value in each bin (plus omni)
- `MM.nRVX` [scalar] number of points at which return value has been computed
- `MM.MarginType` [string] distribution used to transform to standard margins

Since suitable exceedance thresholds are inherently difficult to specify; we recommend the user starts with a wide range for `NEP`, say $[0.3, 0.95]$, working down to a narrower band of thresholds based on Figure 13 (more on this process below).
Note that empty bins will still be assigned GP parameters (in the composite likelihood the empty bin will contribute no information but global values will result).

Output Figures
The following figures illustrate the result of PPC model fitting for the North Sea example on the Hs margin.
The leftmost panel of Figure 7 shows the original response data plotted against covariate $\boldsymbol{X}$. The blue lines represent a 95% confidence interval on the location of the threshold and incorporate two sources of randomness originating from bootstrap re-sampling and from drawing non-exceedance probability τ at random from a uniform distribution over range `NEP`. The solid blue line indicates the threshold used for the original (not bootstrap re-sampled) dataset with τ taken to be the median of all NEPs sampled in range `NEP`. The central and rightmost panels illustrate the transformation of the original dataset to uniform and then Gumbel margins (the process followed using the PIT).
KEY POINT: Any inhomogeneity with respect to direction in the central plot in Figure 7 suggests that the marginal model has not fitted well. In this case, you should adjust the bin-edges (and possibly NEP) to improve your representation of non-stationarity with respect to the covariate(s).
Figures 8 and 9 include 95% confidence intervals on the non-stationary GP scale and stationary GP shape parameters respectively, as a function of the covariate. Again, these are based on bootstrap resampling uncertainty and NEP sampling uncertainty.
Figure 10 illustrates the cross-validation on roughness penalty λ, via a lack-of-fit plot for values within range [`CV.SmthLB`, `CV.SmthUB`].
KEY POINT: If the red line, indicating the optimal choice of λ is at the left or rightmost edge of Figure 10; we have not considered a wide-enough range of roughness penalty values. In this case the range of penalty values considered should be widened by adjusting input `CV.SmthLB` or `CV.SmthUB`.
KEY POINT: The quality of model fit within each covariate bin can be assessed using Figure 11. Red dots outside the confidence limits of the model (plotted in black) indicate a poor fit, in which case the user might reconsider their bin choice and/or NEP range etc.
Empty bins (after thresholding) are indicated by an empty plot window for the associated sector. Figure 12 illustrates the overall goodness of fit.
KEY POINT: Figure 13 is a key output plot, showing how the estimated GP shape parameter varies as a function of the non-exceedance probability. This plot should be used to narrow down on an NEP range over which the GP shape is relatively stable. In the left panel a reasonable choice might be [0.5,0.75], in the right panel a narrower range might be chosen, say [0.5, 0.65]. The right limit can usually be chosen as the last point before which the confidence interval widens or there is a change in slope. The lower limit should generally not be below the mode of the data since we are fitting a tail model. We choose to use an ensemble of thresholds in our analysis in recognition of the challenge of threshold selection in extreme value statistics.
Finally, Figure 14 provides 10 and 100 year return level cumulative distribution functions for each covariate (here, directional) sector. When there are fewer colours in the plot than the legend; one or more of the CDFs overlap. Empty sectors are listed in the legend with an "Empty Bin" description and do not have an associated CDF curve.

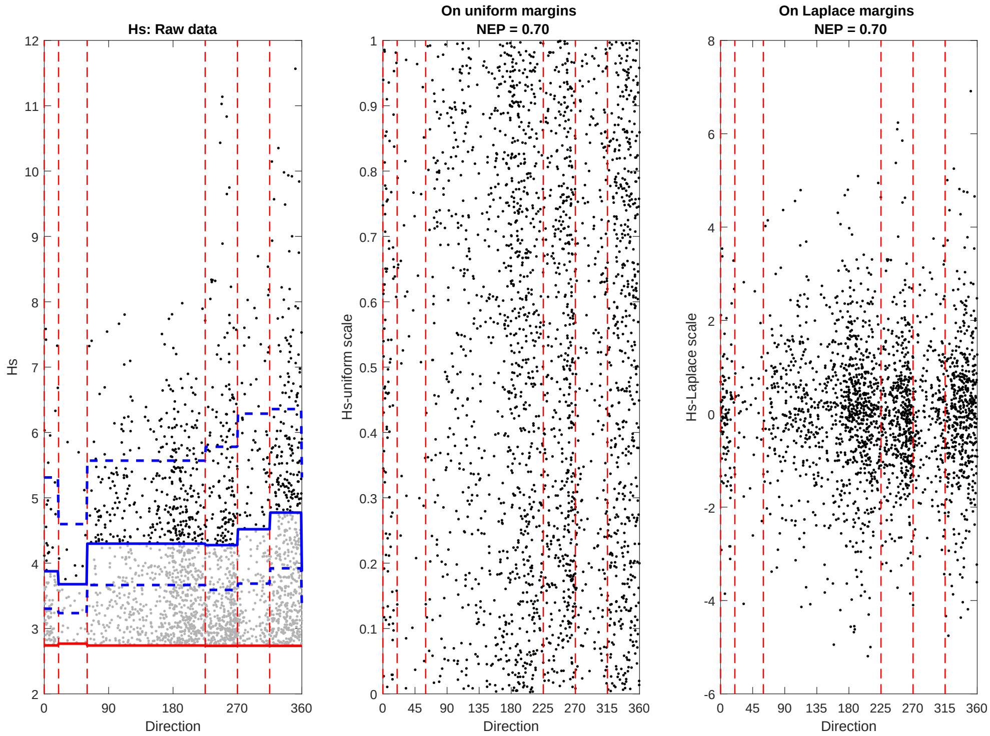

Figure 7: Left panel shows sea-state data in grey and storm peaks in black. Bin edges are indicated by dashed red lines. 2.5, 50 and 97.5 percentiles of estimated threshold across all bootstraps are plotted with blue lines. The Gamma location parameter is plotted in solid red. Middle and right panels show data transformed to uniform and Laplace scale.

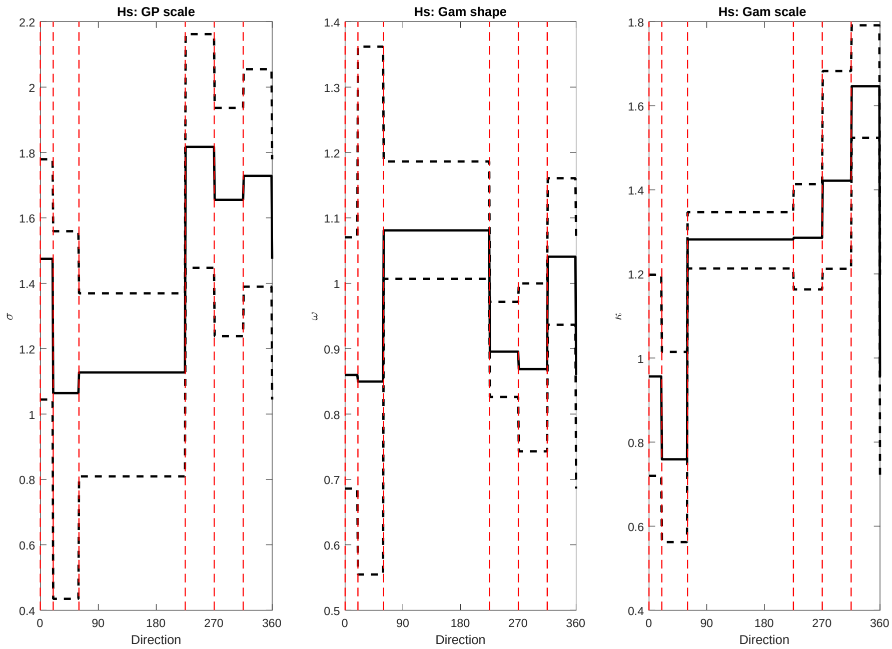

Figure 8: Black lines show 2.5, 50 and 97.5 percentiles of GP shape as a function of direction. Red lines show bin edges.

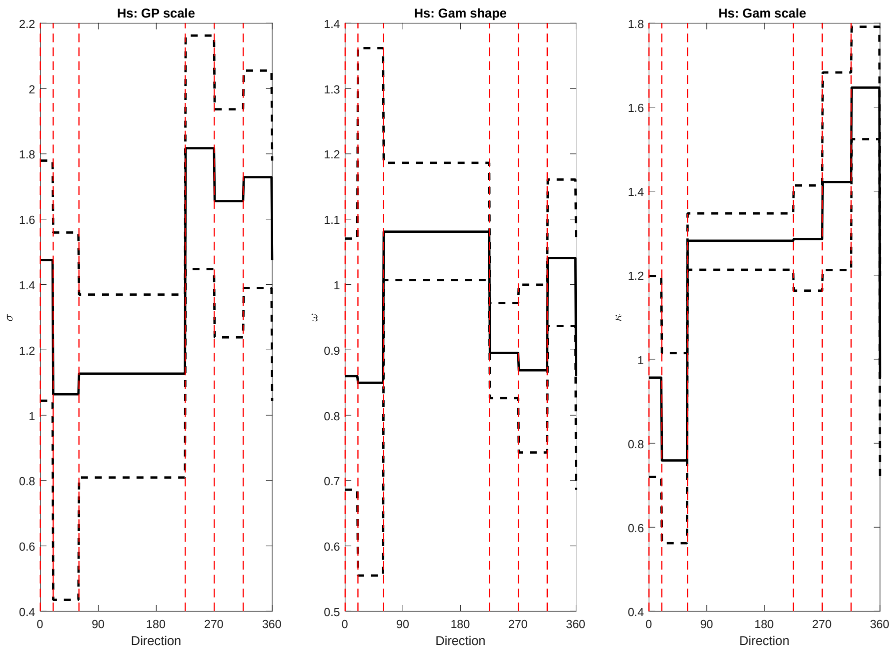

Figure 9: Black lines show 2.5, 50 and 97.5 percentiles of GP shape as a function of direction. Shape parameter is constant w.r.t to covariate.

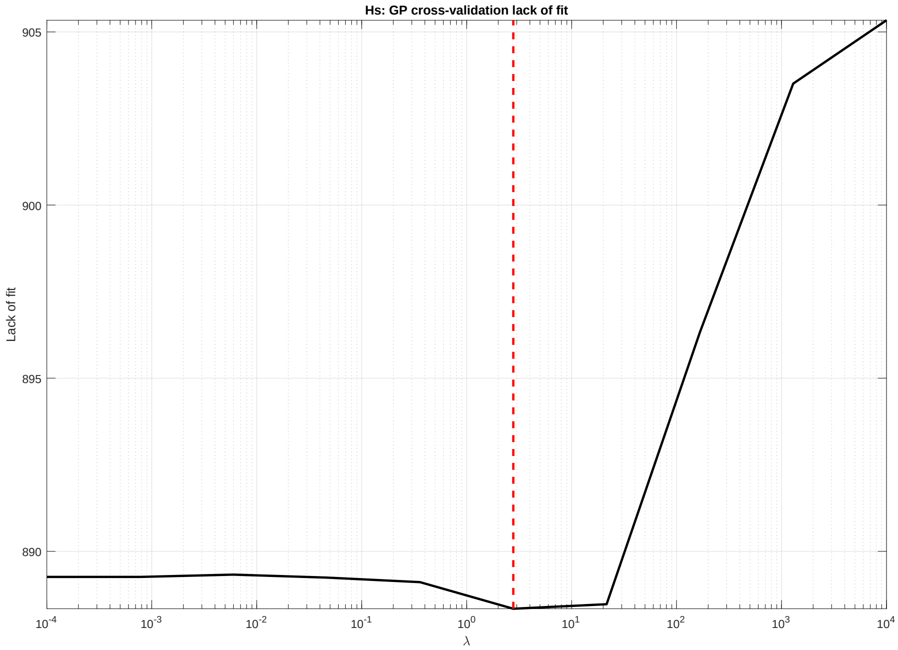

Figure 10: Cross Validation plot showing lack of fit against chosen smoothness λ of GP scale. Low indicates good prediction performance. The red line indicates the optimal choice.

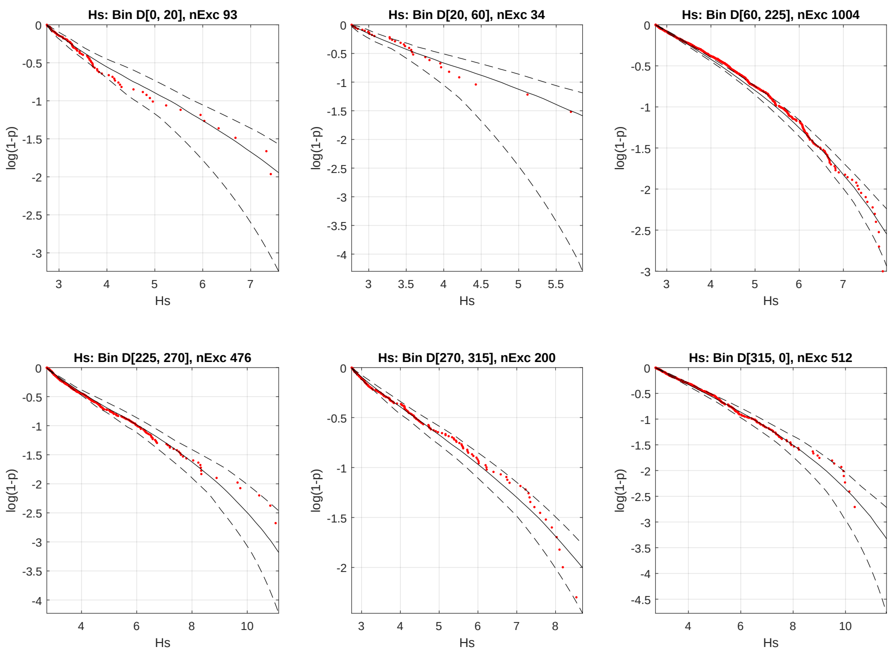

Figure 11: Diagnostic for quality of model-fit, broken out by covariate (here, directional) sector. Red dots show storm peaks, black lines are 2.5, 50 and 97.5 percentiles of model prediction over bootstraps. Red dots inside the confidence limits of the model indicate good fit.

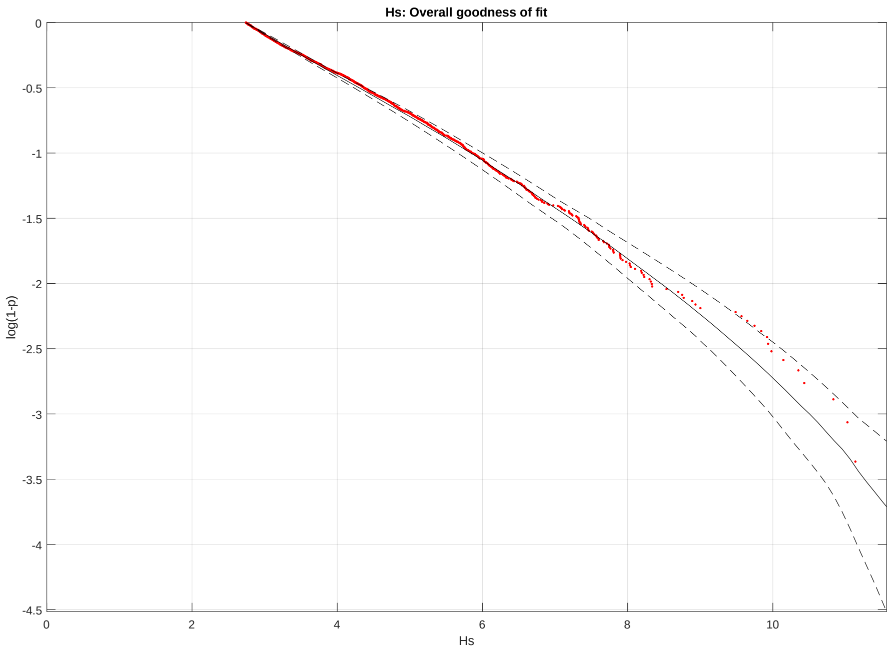

Figure 12: Diagnostic for overall quality of model-fit. Red dots show storm peaks, black lines are 2.5, 50 and 97.5 percentiles of model prediction over bootstraps. Red dots inside the confidence limits of the model indicate good fit.

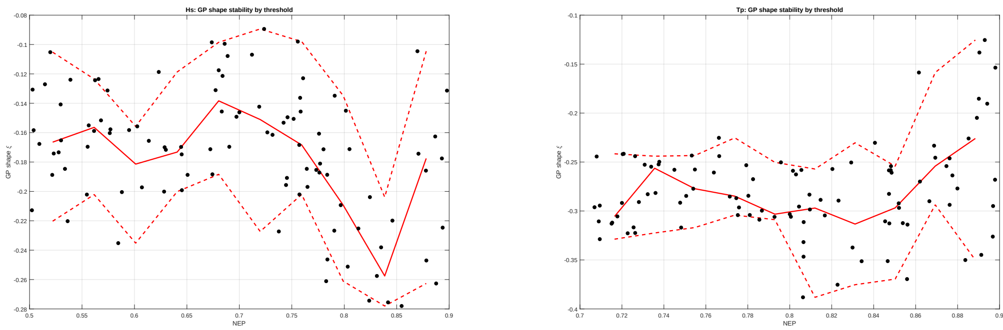

Figure 13: GP shape ξ as a function of the NEP for 2 responses Hs (left plot) and Tp (right plot) from the North Sea data. Black dots show individual bootstrap estimates, red lines are local binned median, 2.5 and 97.5 percentile estimates. A well behaved model should be stable over a range of NEP's.

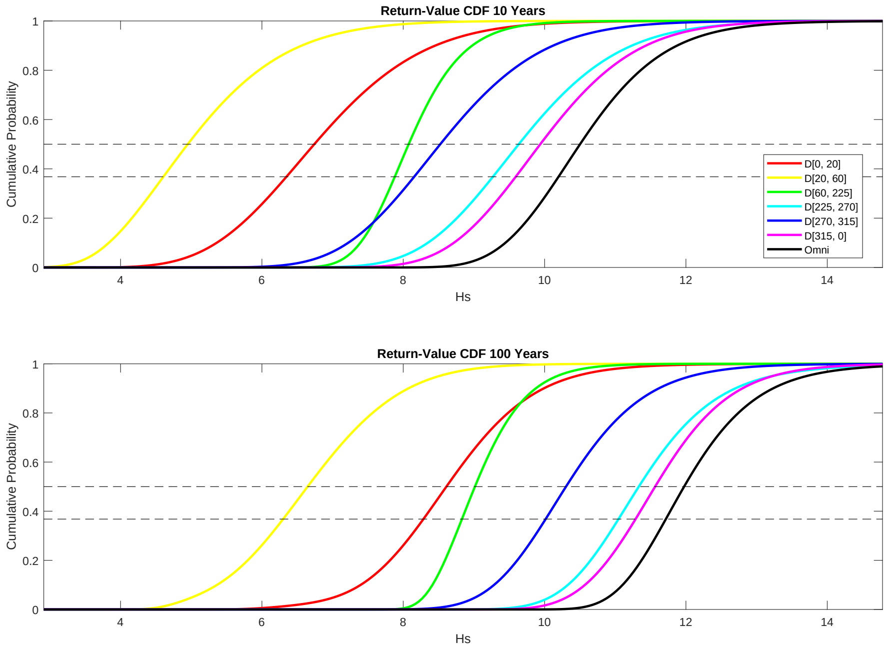

Figure 14: Marginal 10 year (upper plot) and 100 year (lower plot) return value CDFs (Hs). Directional sectors are show using coloured lines. The black line shows the omni-directional estimate.

4 Stage 4: Fit conditional extremes model

4.1 Conditional extremes model

The standard-scale (Gumbel or Laplace) sample $\{y_{1i}, y_{2i} \ldots, y_{Di}\}_{i=1}^{N}$ above some threshold of the conditioning variate Y_1 is used to estimate a conditional extremes model with parameters $\boldsymbol{\alpha}_{\tilde{\tau}k}, \boldsymbol{\beta}_{\tilde{\tau}}, \boldsymbol{\mu}_{\tilde{\tau}}$ and $\boldsymbol{\sigma}_{\tilde{\tau}}$

$$(Y_2, Y_3 \ldots Y_D | Y_1 = y_{i1}) = \boldsymbol{\alpha}_{\tilde{\tau}k} y_{i1} + y_{i1}^{\boldsymbol{\beta}_{\tilde{\tau}}} \boldsymbol{W}_{\tilde{\tau}} \text{ for } y > \phi_{\tilde{\tau}k},$$

where $\boldsymbol{W}_{\tilde{\tau}} \sim N(\boldsymbol{\mu}_{\tilde{\tau}}, \boldsymbol{\sigma}_{\tilde{\tau}}^2)$ is assumed for model estimation only. Threshold $\phi_{\tilde{\tau}k}$ is defined as the quantile of the standard Gumbel distribution with non-exceedance probability $\tilde{\tau}$ for covariate bin C_k. Note that there the non-exceedance probability for the HT model need not be the same as the marginal model.

4.2 Running Stage4

Running Stage4 fits a Hefferenan and Tawn conditional extreme value model. Marginal model data and parameters (`Output\MM1.mat` and `Output\MM2.mat`) are loaded from stage 3, described in section 3.

Run Scripts: `Stage4_FitHeffernanTawn.m`
Output files: `Output\HT.mat`
Figures files:

- `Figures\Stg4_HT_1_SmlvsData`
- `Figures\Stg4_HT_2_ResidualDiagnostics`
- `Figures\Stg4_HT_3_Parameters`
- `Figures\Stg4_HT_4_AlphaThresholdStability`
- `Figures\Stg4_HT_4_BetaThresholdStability`
- `Figures\Stg4_HT_6_ConditionalReturnValueCDF`

Inputs:

- `HTNEP` = Conditional non exceedence probability range. Make sure HT NEP$>\exp(-\exp(-0)) = 0.368$ or the Gumbel transformation will fail.
- `NonStationary` = 0: use a stationary α parameter in HT; 1: fit penalised piecewise constant (non-stationary) α using the same bins as the marginal analysis
- `CV.CVMth` = 0: Only cross validate roughness penalty for original dataset (fast); or 1: Cross Validate smoothness for every bootstrap re-sample (slow)
- `CV.nCV` = number of cross validation groups
- `CV.nSmth` = number of roughnesses tried in CV
- `CV.SmthLB` = lower bound (log10) for roughness range
- `CV.SmthUB` = upper bound (log10) for roughness range
- `SampleLocalResid` = If this is set to true, when simulating under H&T model residuals are re-sampled locally (from the current covariate bin); if false, residuals are sampled globally, i.e. from any bin.

Outputs: in file `HT.mat`

- `HT.Prm` $[nPrm \times (D-1) \times nB]$ HT model parameters
- `HT.Rsd` $[nB \times 1]$ cell array sampled residuals from each bootstrap

- HT.Thr $[nB \times D-1]$ HT threshold used
- HT.NEP $[nB \times 1]$ non-exceedence probabilitys
- HT.nB [scalar] number of bootstraps
- HT.X $[N \times (D-1) \times nB]$ transformed conditioning variable on standard scale
- HT.Y $[N \times (D-1) \times nB]$ transformed associated variable on standard scale
- HT.RV.X_Stn: $[(nBin+1) \times nRls \times R]$ simulated return values for conditioning value on standard scale
- HT.RV.X: $[(nBin+1) \times nRls \times R]$ simulated return values for conditioning value on original scale
- HT.RV.Y_Stn: $[(nBin+1) \times (D-1) \times nRls \times R]$ simulated associated return values for conditioning value on standard scale
- HT.RV.Y: $[(nBin+1) \times (D-1) \times nRls \times R]$ simulated associated return values for conditioning value on original scale
- HT.RV.nRls [scalar] number of realisations used in conditional return value simulation
- HT.n number of observations N
- HT.nDmn [scalar] number of dimensions D
- HT.SmpLclRsdOn [boolean] flag for sampling for residual from local bin
- HT.nAlp [scalar] number of alpha parameters in the model
- HT.nPrm [scalar] number of model parameters
- HT.nBin [scalar] number of covariate bins
- HT.nRtr [scalar] number of return periods
- HT.NonStat [boolean] non stationary alpha parameter flag
- HT.A $[N \times nB]$ bin allocation
- HT.RsdInd $[nB \times 1]$ cell [sampled residuals indexes
- HT.CVMth [boolean] cross validation method
- HT.nCV [scalar] number of Cross validation groups
- HT.nSmth [scalar] number of smoothness parameters used
- HT.SmthLB [scalar] lower bound for roughness set
- HT.SmthUB [scalar] upper bound for roughness set
- HT.SmthSet $[1 \times nSmth]$ [set of roughness parameters]
- HT.OptSmth $[1 \times nB]$ Optimal smoothness chosen
- HT.CVLackOfFit $[nSmth \times nB]$ Lack of fit for roughness estimation
- HT.MarginType [string] margin type for transforming to standard scale

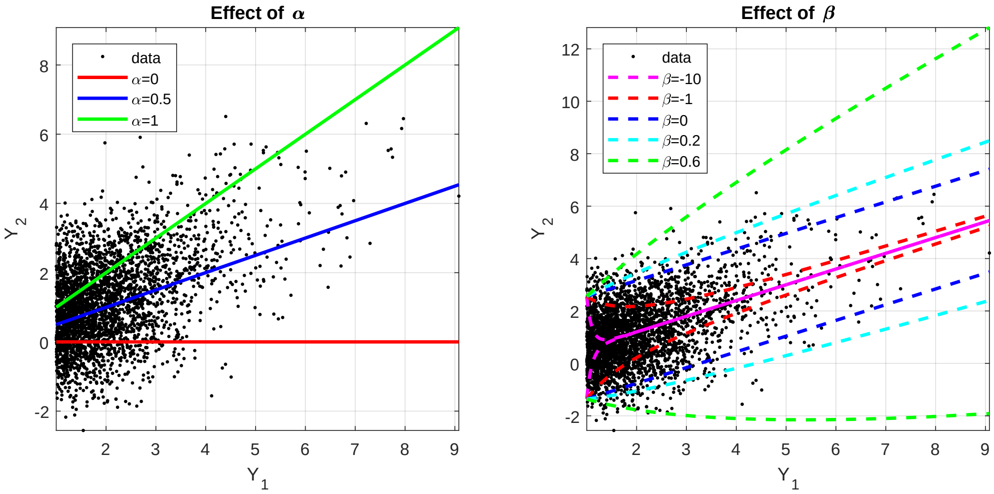

Figure 15: Illustration of the impact of HT parameters α and β on the structure of dependence between two Gumbel distributed random variables

Note that sampling residuals locally (setting `SampleLocalResid` = true) effectively gives a non-stationarity to the residual part of the H&T model, thus improving the fit/simulation procedure. That said, in the presence of bins with very few observations, we advice that this functionality is turned off (set to false) as the simulated data for a bin with very few observations will come from resampling a very small set of residuals many times. In this case it is therefore better to sample globally to increase the number of residuals from which the simulation resamples.

The number of bootstrap re-samples is inherited from the marginal model settings. Plots of model parameter estimates and residual distributions as a function of threshold and covariates aid selection of an interval of thresholds consistent with modelling assumptions. The conditional extremes model form is $(Y_2, Y_3, \dots | Y_1 = y) = \boldsymbol{\alpha} y + y^{\boldsymbol{\beta}} W$ for $y > \phi$, for sufficiently large threshold ϕ, where $\boldsymbol{\alpha} \in [0,1]$ and $\boldsymbol{\beta} \in (-\infty, 1]$. W is a random variable with an unknown distribution, the density of which we estimate using residuals from the fitted model. For fitting purposes only, we assume that $W \sim N(\boldsymbol{\mu}, \boldsymbol{\sigma}^2)$. The model fitting corresponds to estimating $\{\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\mu}, \boldsymbol{\sigma}\}$ given a sample of values for $\{Y_1, Y_2, \dots\}$. All of $\phi, \boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\mu}$ and $\boldsymbol{\sigma}$ are in principle functions of covariates. Using the conditional extremes model, we simulate joint extremes on the standard Gumbel scale, and transform these realisations to the original scale using the probability integral transform once more. Below the threshold ϕ, realisations can be simply re-sampled from the sub-sample of the original data for which $y \leqslant \phi$.

Figure 16 shows a comparison of the data and a simulation from the HT model. On the original scale, two different spikes can be seen in the upper right hand tail of the joint distribution, reflecting different marginal characteristics in T_P. Figure 17 shows the results on the North Sea $\mathrm{T}_P|\mathrm{H}_S$ example. It is typical that these residuals are quite skewed (not normal), which is why they are reused in the simulation procedure. Figure 18 shows model parameter $\boldsymbol{\alpha}$ for the stationary case is near 1, this indicates strong dependency between large H_S and T_P. In the non-stationary case, Figure 19 shows $\boldsymbol{\alpha}$ fairly similar in most sectors but it is highly uncertain in the sector where there is no data. $\boldsymbol{\alpha}$ nearer 0 would indicate weak or no dependency. Figure 15 illustrates the influence of $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ parameters on the shape of dependence.

The threshold stability plots in Figures 20 and 21 are similar to those in Figure 13. These should be used in the same way as described in section 3.3 to find a suitable range for the HT NEP. A range of [0.5,0.7] would seem to be a reasonable choice here. Figure 22 shows the return value CDFs for the North Sea $\mathrm{T}_P|\mathrm{H}_S$ example. Here the omni directional CDF is bimodal, this is largely due to directional differences in the T_P marginal distribution. Similar effects can be seen in Figure 16.

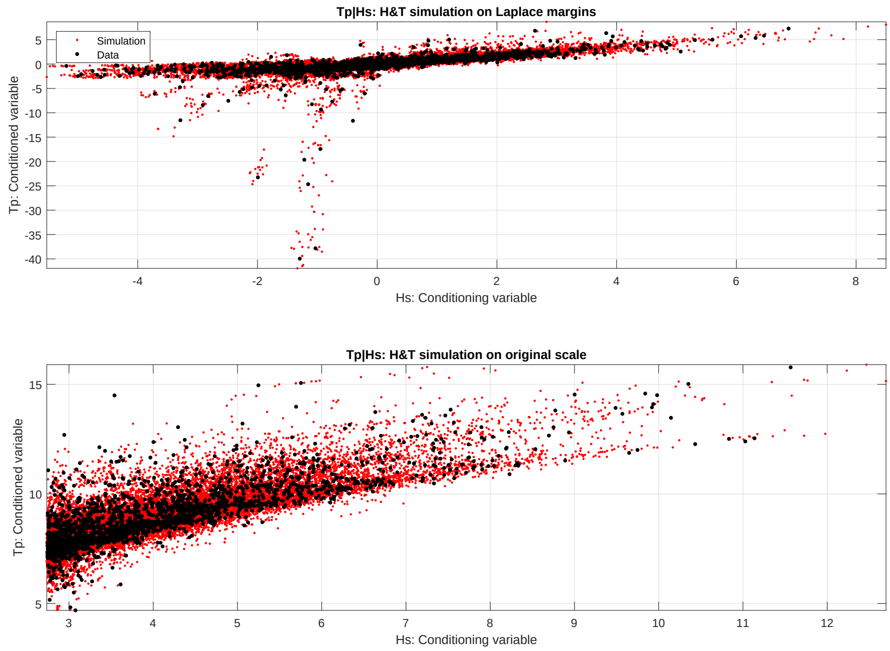

Figure 16: Comparison of original data (black) and simulation from fitted H&T model for 10× period of the data (red) on standard margins (upper plot) and original margins (bottom plot). On the original scale two different spikes can be seen at the upper right corner reflecting different marginal characteristics in T_P

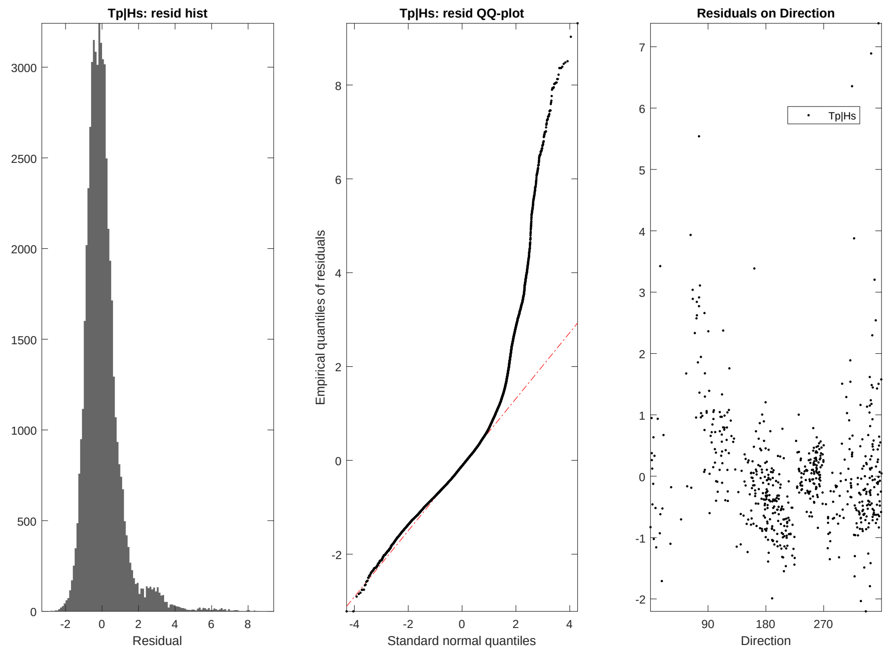

Figure 17: Diagnostic of the residuals from the HT fitting. Left panel shows a histogram of the residuals. Middle panel shows a normal QQ plot. The right panel shows residuals as a function of direction. It is typical that these residuals are quite skewed (not normal), which is why they are reused in the simulation procedure.

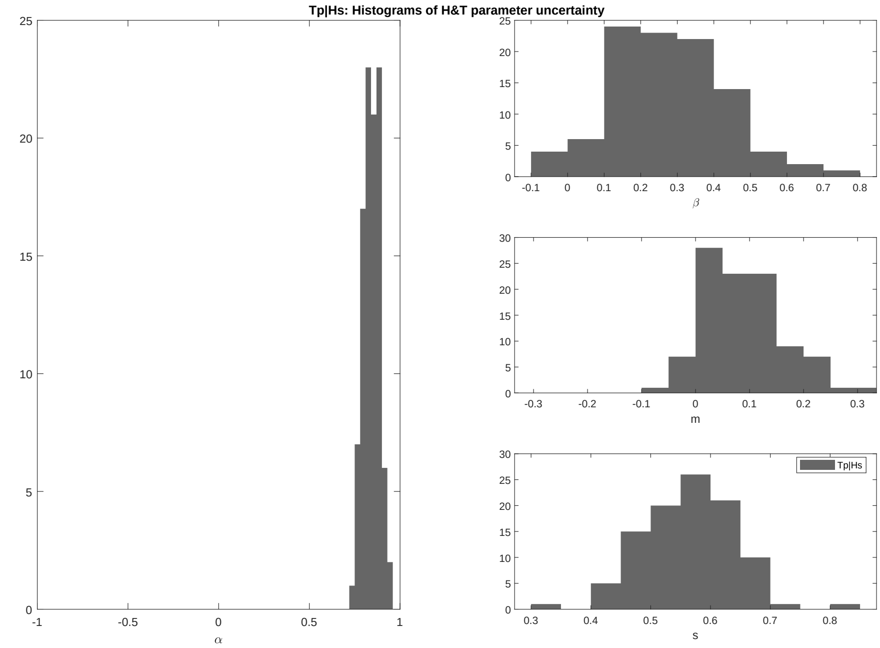

Figure 18: Histograms of the HT parameters over bootstrap re-samples in a stationary case. The parameter uncertainty captures, marginal (bootstrap and NEP) and conditional (bootstrap and HT NEP) uncertainty.

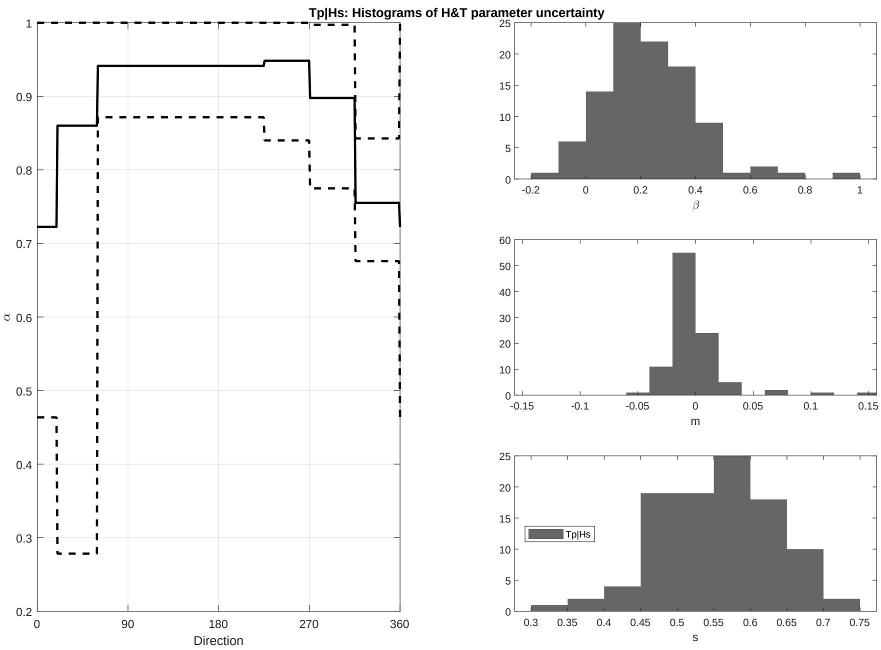

Figure 19: Histograms of the HT parameters over bootstrap re-samples in a non-stationary case. The parameter uncertainty captures, marginal (bootstrap and NEP) and conditional (bootstrap and HT NEP) uncertainty.

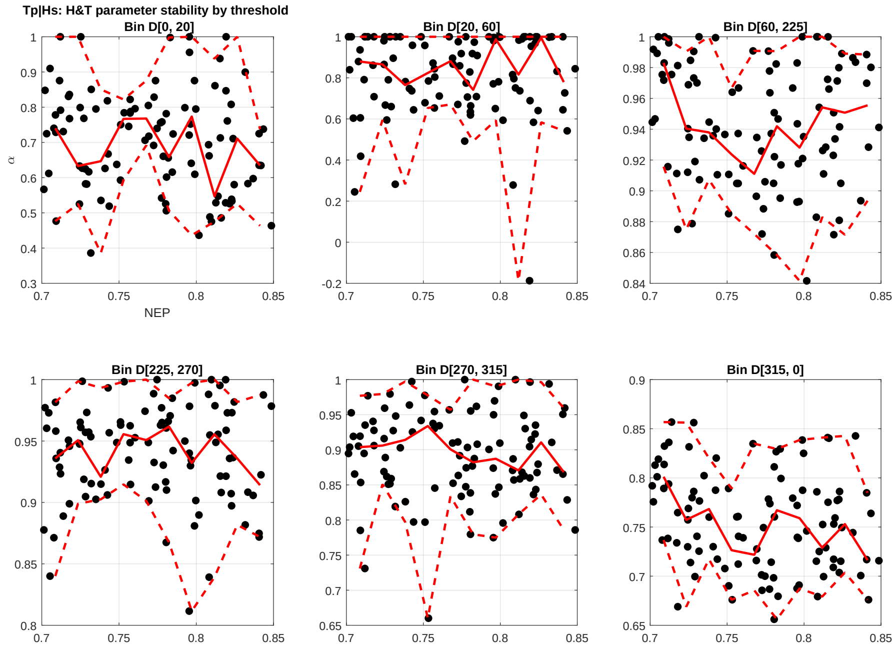

Figure 20: HT parameter α as a function of the HT NEP. Black dots show individual bootstrap estimates, red lines are local binned median, 2.5 and 97.5 percentile estimates. A well behaved model should be stable over a range of NEP's

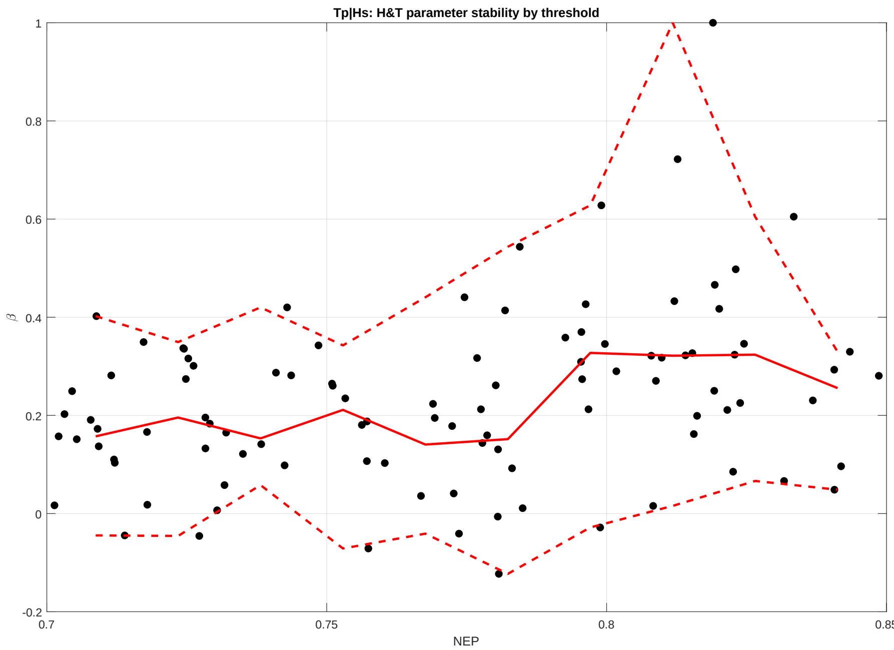

Figure 21: HT parameter β as a function of the HT NEP. Black dots show individual bootstrap estimates, red lines are local binned median, 2.5 and 97.5 percentile estimates. A well behaved model should be stable over a range of NEP's

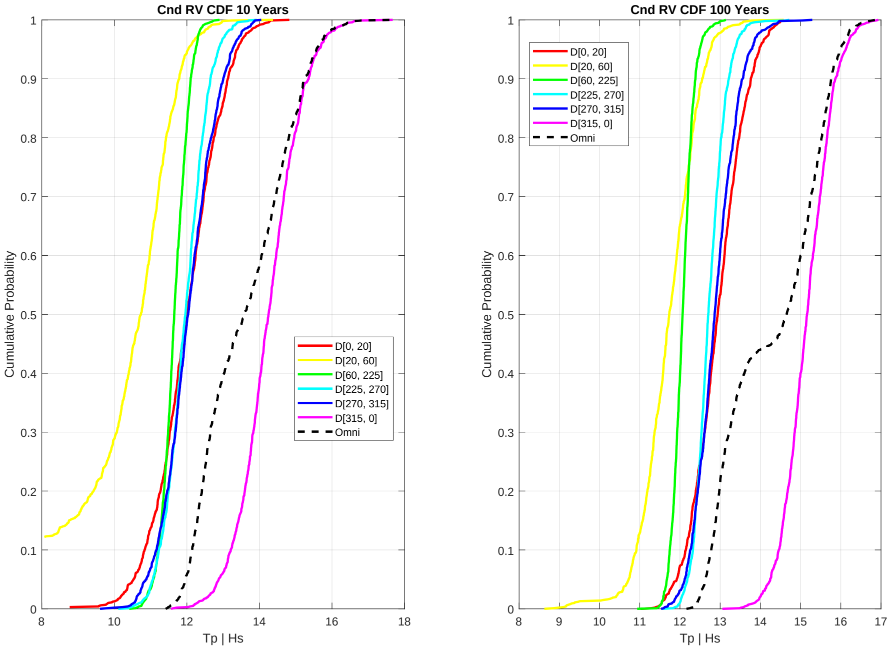

Figure 22: Conditional return value CDF's $p(\mathrm{Tp}|Hs_{10})$ and $p(\mathrm{Tp}|Hs_{100})$. Directional CDFs are shown using coloured lines. Black line shows the omni-directional estimate. In this case using the North Sea data the omni directional CDF is bimodal, this is largely due to marginal differences in Tp. Similar effects can be seen in figure 16

5 Stage 5: Draw contours

Three approaches are used to estimate extreme contours using the marginal and H&T models: constant exceedance; direct sampling (Huseby) and Heffernan & Tawn density contours. These methods are described briefly below. The reader is directed to Haselsteiner et al. [2017] for an excellent recent review of contour methods, and [Ross et al., 2019] (output of the ECSADES project) who discuss best-practice in the application of contours.
All the contours pass through a lock point, defined using the extreme quantile in Y_1 and the conditional median in Y_2. A new method for computing contours `Sml=SimulateIS(HT,Mrg,nRls)` is used to sample uniformly across possible range of model values used for importance sampling of contours Importance reweighting of these simulated values then gives smooth contours at all return periods using far less simulations from the HT model.

5.0.1 Constant Exceedance Contour

Preserve the observation count in the extremal set, as illustrated in Figure 25. We have chosen to use a quadrant, this is of course arbitrary and many different shapes could be considered e.g tangent, half plane, etc.

- Downside: convexity means contour will never come back on itself - pushing the upper part of the curve into a table top. The resulting curve is not very practical.
- Upside: definition is sound in the probabilistic sense, none of the others preserve probabilities. From a risk point of view it's robust.

5.0.2 Huseby Contour

A convex contour based on the work of Huseby et al. [2015]. This is similar to the constant exceedance contour except that a tangential set is used instead of the quadrant. The contour is computed in angular space around a centre point and fast changing parts of the contour can be a little spiky. A moving average is used to try to overcome some of these issues but this does not always work.

- Downside: The convexity doesn't behave well in multimodel cases or other cases where the data is non-convex.
- Upside: A complete contour is produced for all directions not just the largest sets.

5.0.3 Hefferenan and Tawn density

Grid the data on the original scale and work out density of simulations in each bin. Contour is defined by preserving the density to that seen in the bin containing the lock point. Figure 26 illustrates this method. The omni contour is computed as a weighted sum across the binned contours (weighted by rate of occurrence).

- Downside: the contour is non-invariant to transformations of variables
- Upside: we get a contour more like what we'd expect (without the table top).

5.1 Running Stage5

Marginal and conditional model parameters are loaded from Stage 3 and Stage 4 respectively.
Run Scripts: `Stage5_Contour.m`
Output files: `Output\Cnt.mat`
Figures files:

- `Figures\Stg5_Contour_1_Omni`
- `Figures\Stg5_Contour_2_Binned_#`

Inputs:

- `Mth` = cell array of contour methods to be used:
  - `Exc`: constant exceedance contour
  - `HTDns`: constant density contour on standard margins, uses density form of H&T to get contour
  - `Hus`: convex "Direct Sampling" contour of Huseby
- `nSml` = number of simulations under H&T model (upon which contours are estimated). May need to increase this when you have lots of bins, or see lack of smoothness in e.g. Huseby contour

Output: Contour structure `Cnt`

- `Cnt.nPon` [scalar] how many points from which to draw contour
- `Cnt.Rng` $[nPon \times (nBin + 1) \times Cnt.nLvl]$ conditioned values for contour
- `Cnt.XY` $[nMth \times 1]$ cell array for the contour lines. In case of Exc and Hus methods $XY(i)$ is $[nPon \times 2 \times (nBin + 1) \times R \times (D - 1)]$ defining contour lines in case of HTDns $XY(i)$ is a $[(nBin + 1) \times R \times (D - 1))]$ cell with sub-elements $[2 \times nPon]$ defining the contour in this case nPon varies for each contour bin, return period and associated variable.
- `Cnt.Mth` $[nMth \times 1]$, (cell array) of contour methods used
- `Cnt.nMth`[scalar], number of contouring methods
- `Cnt.nBin`[scalar], number of covariate bins
- `Cnt.nLvl`[scalar], number of contour levels chosen
- `Cnt.nAsc`[scalar] number of associated variables
- `Cnt.Sml` structure importance sampled simulation under the model
- `Cnt.LvlOrg` $[(nBin + 1) \times R \times (D - 1)]$, contour level on original scale of conditioned variable (lock point)

Figures 23 and 24 show comparisons of the 3 methods for the North Sea $T_P|H_S$ example.

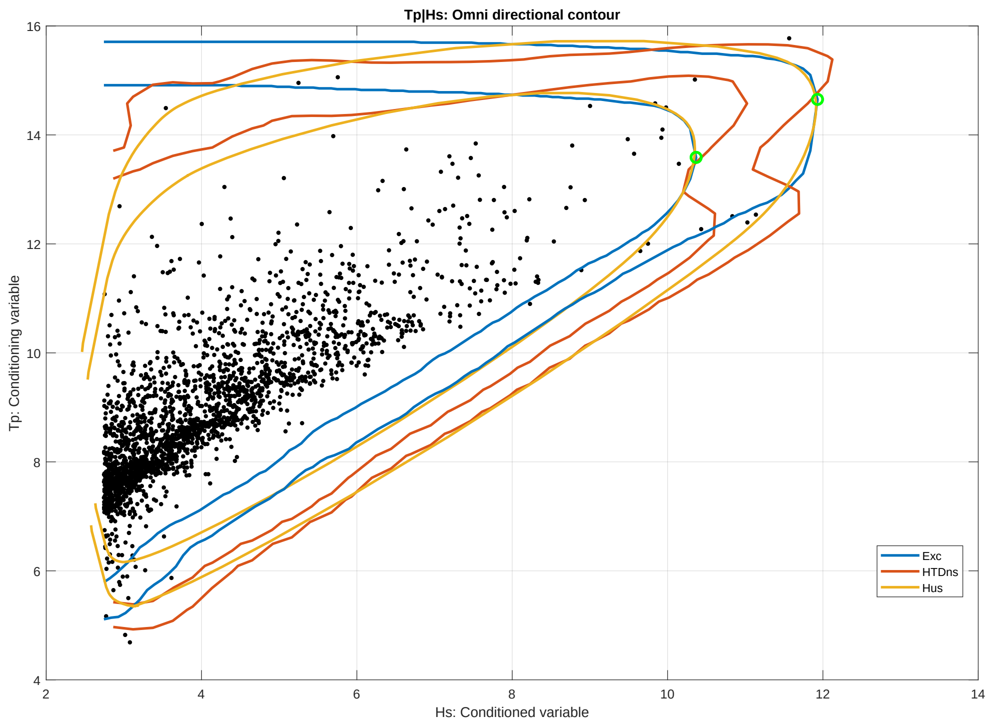

Figure 23: Comparison of contour methods omni directionally. Contours are for 10 and 100 year return periods. Different methods are shown using coloured lines. The software currently does not support an omni HTDns method. The green circle shows the lock point at each probability level through which all the contour methods have to pass.

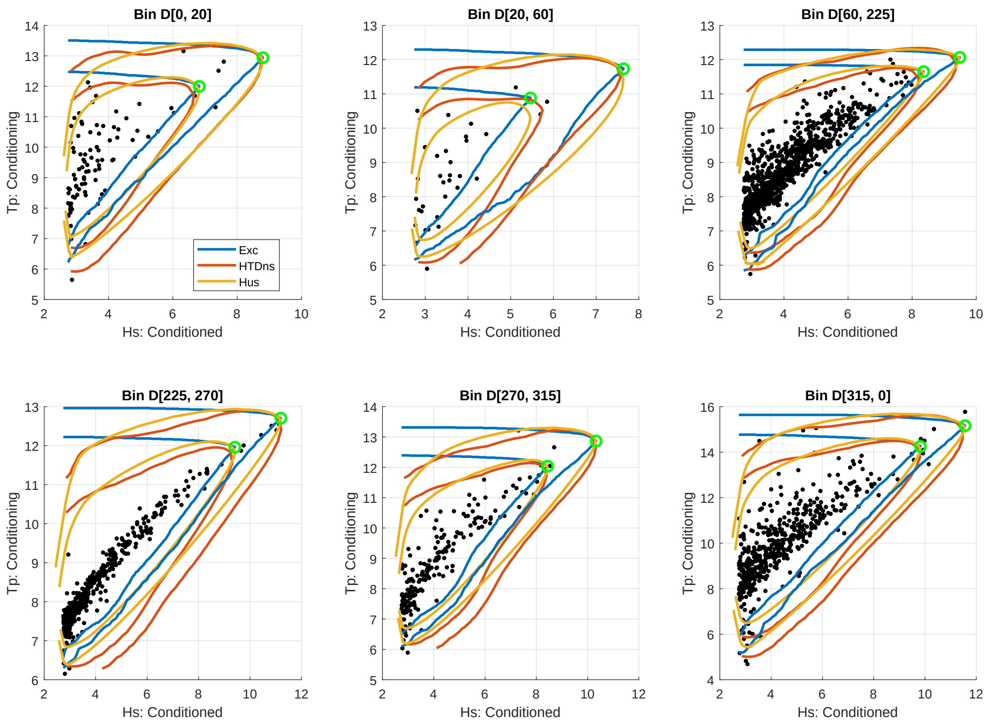

Figure 24: Comparison of contour methods by directional sector. Contours are for 10 and 100 year return periods. Different methods are shown using coloured lines. The green circle shows the lock point at each probability level through which all the contour methods have to pass.

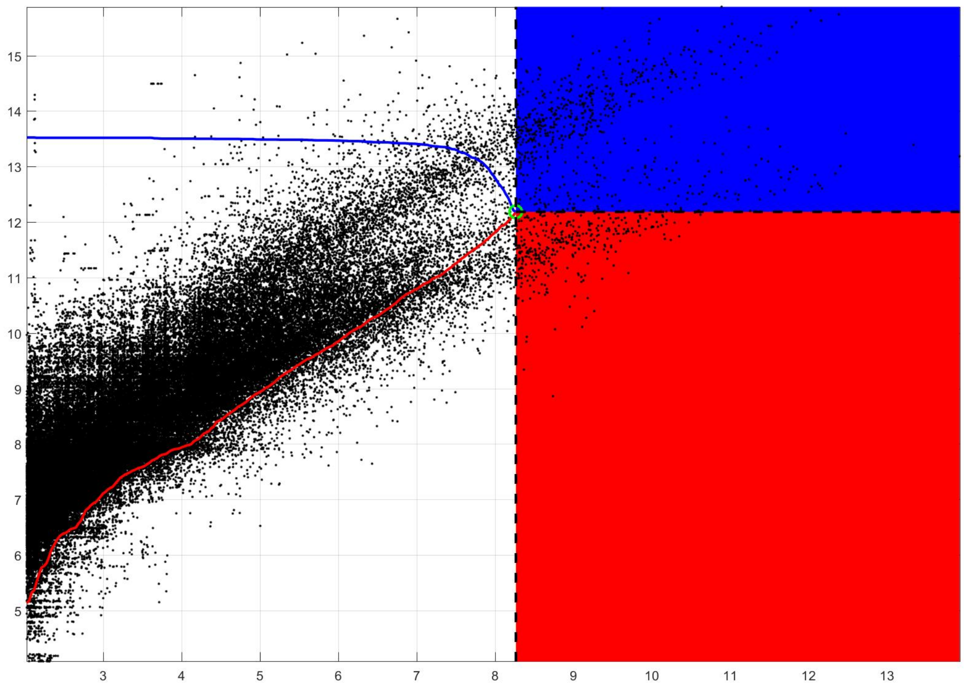

Figure 25: Illustration of constant exceedance contour. Green lock point is defined using the extreme quantile in Y_1 and the conditional median in Y_2. The blue line is drawn such that, as Y_1 is decreased, the number of observations in the blue quadrant is preserved. The red line is drawn such that, as Y_1 decreases, the number of observations in the red quadrant is preserved.

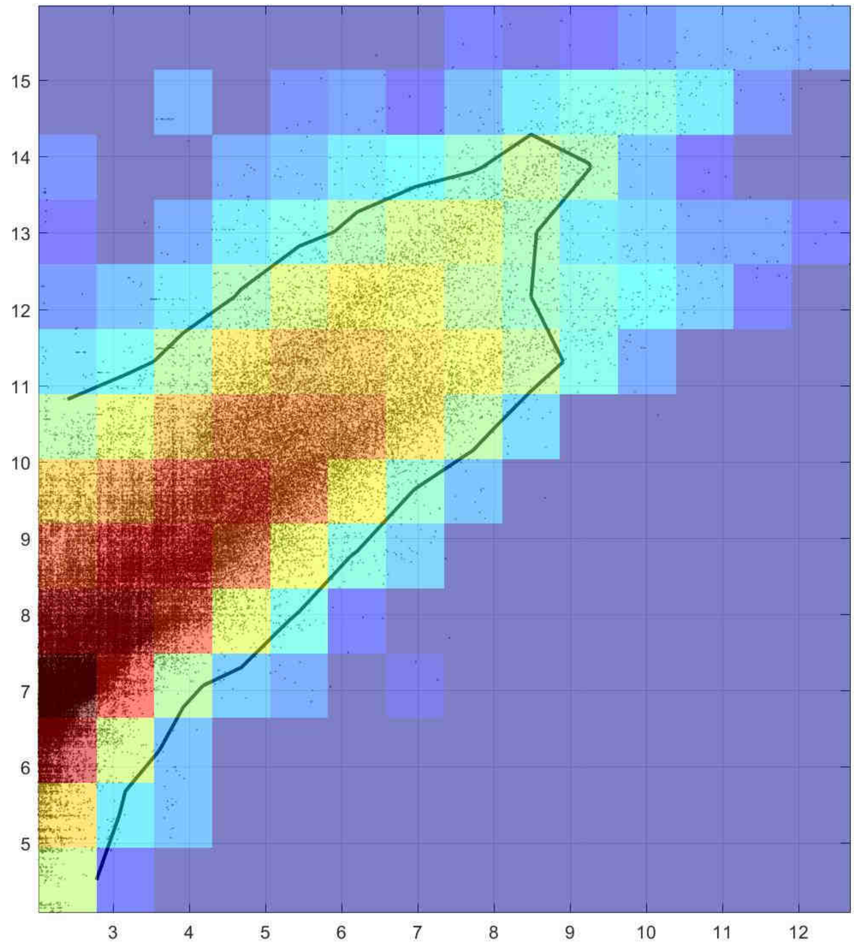

Figure 26: Illustration of empirical density contour. The coloured squares represent the observation counts per bin.

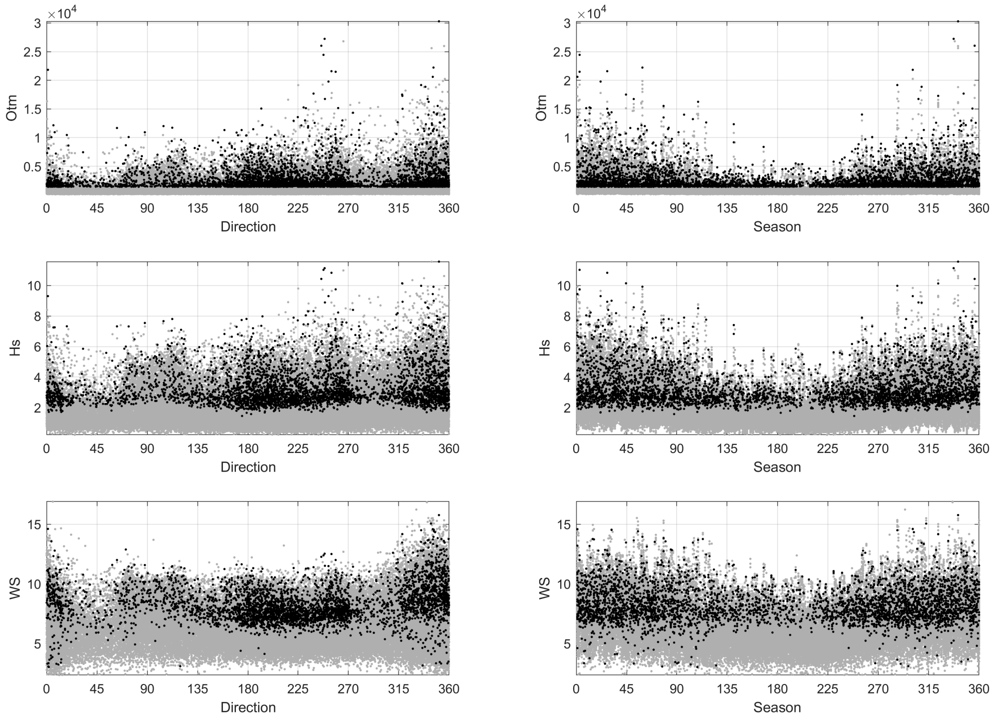

Figure 27: Scatter plots of storm peak responses broken out by bin.

6 Example: Multiple Associated Variables and Covariates

Having illustrated the key inputs and outputs of the code for a simple case - single associated variable and single covariate - in this section we provide a higher-dimensional example. We set the main (conditioning) response to Over-turning Moment and use two associated variables - wind speed and H_S. Further, we will use two covariates: wave direction (periodic) and season, non-periodic as (for illustration) we use day-of-the-year. Figures from the analysis are given below:.

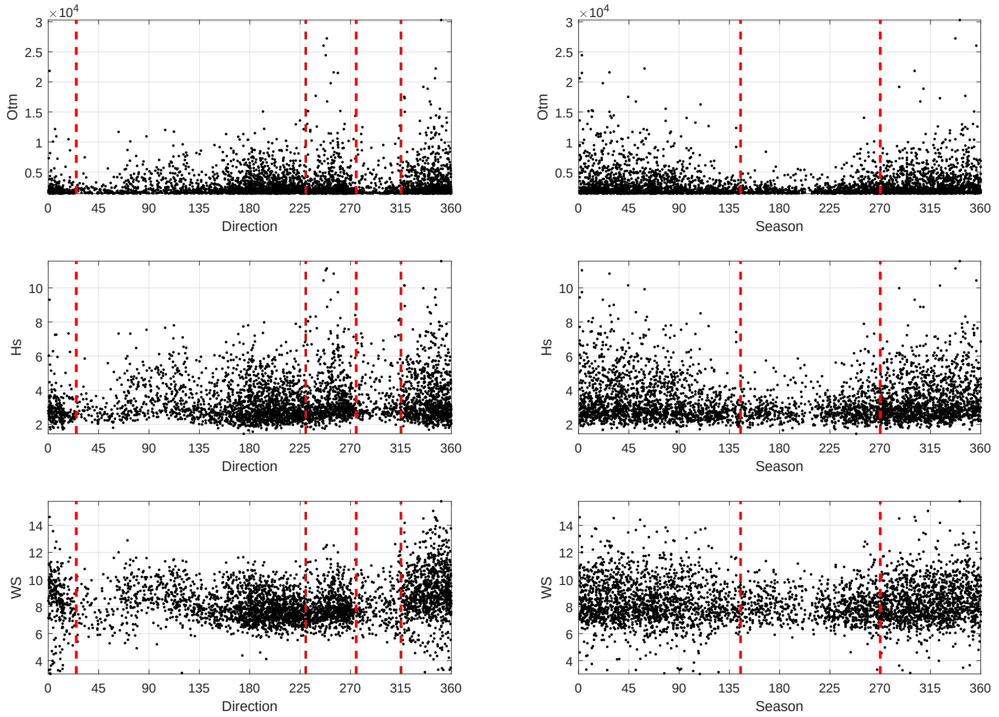

Figure 28: Example bin allocation. Bins are chosen at $[25, 230, 275, 315] \times [145, 270]$ degrees in direction and season. Storm peak data shown in black, chosen bin edges are shown with red dashed lines.

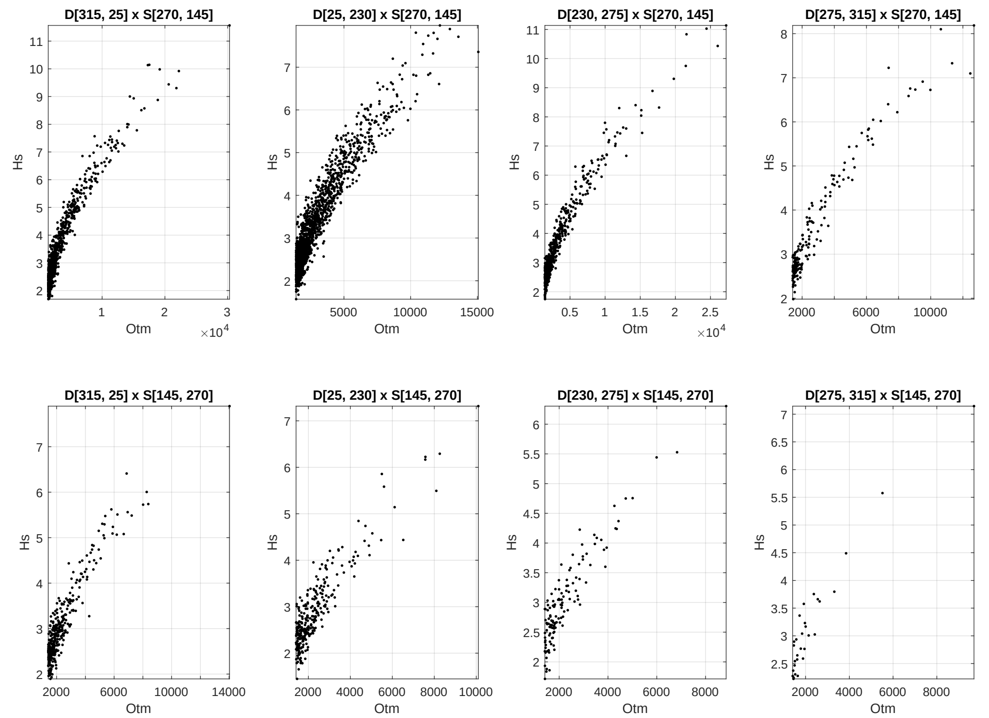

Figure 29: Scatter plots of storm peak responses broken out by bin.

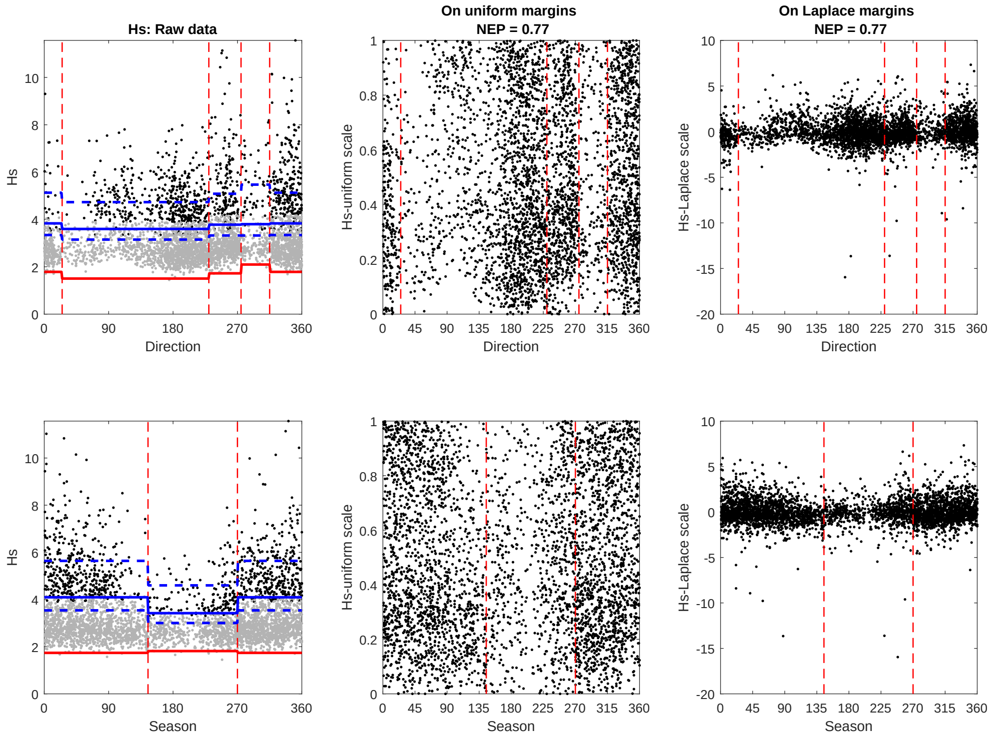

Figure 30: Left column shows sea-state data in grey and storm peaks in black by direction and season. Bin edges are indicated by dashed red lines. 2.5, 50 and 97.5 percentiles of estimated threshold across all bootstraps are plotted with blue lines. The Gamma location parameter is plotted in solid red. Middle and right columns show data transformed to uniform and Laplace scale.

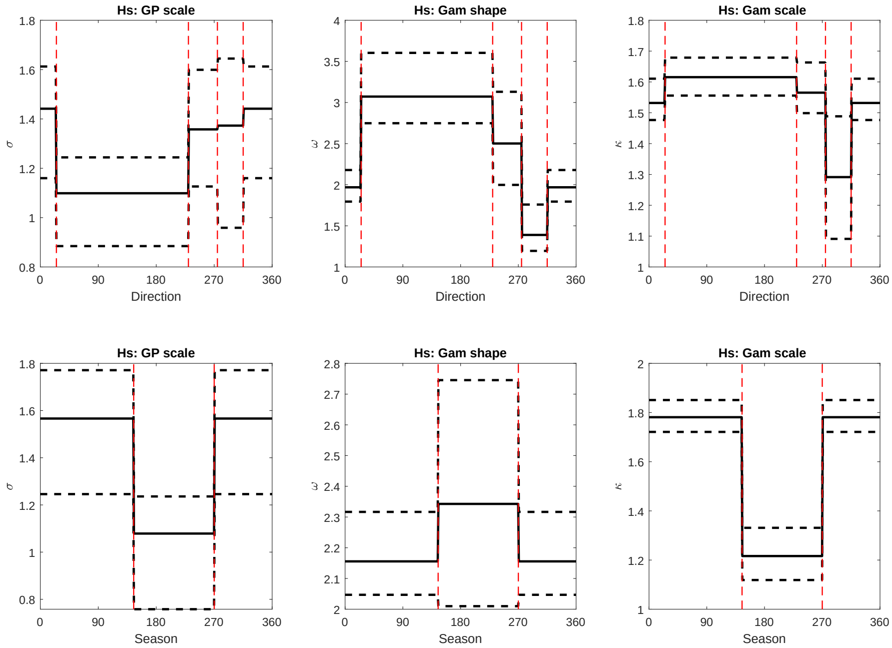

Figure 31: Black lines show 2.5, 50 and 97.5 percentiles of GP scale as a function of direction and season. Red lines show bin edges.

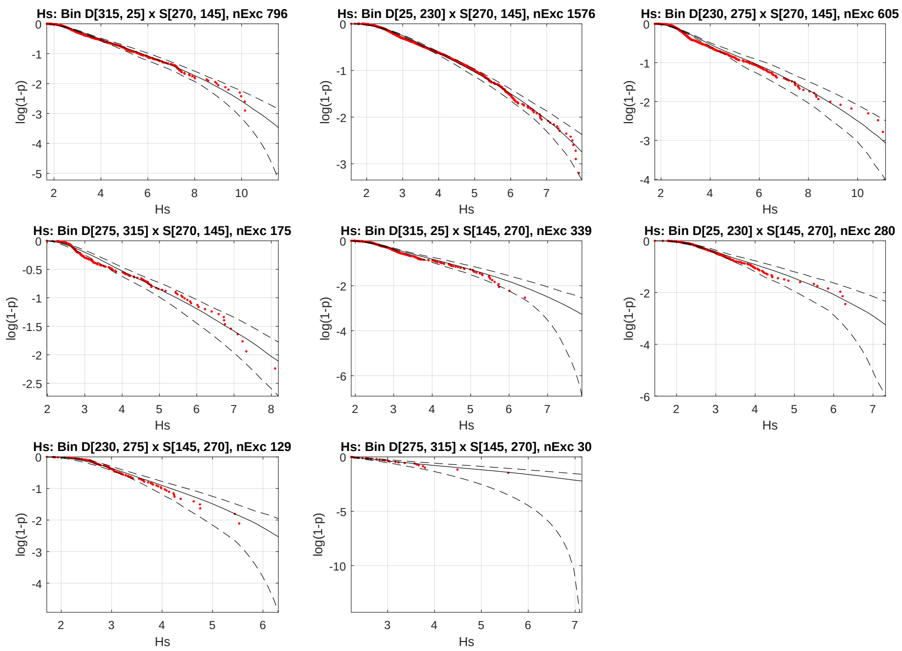

Figure 32: Diagnostic for quality of model-fit, broken out by each sector. Red dots show storm peaks, black lines are 2.5, 50 and 97.5 percentiles of model prediction over bootstraps. Red dots inside the confidence limits of the model indicate good fit.

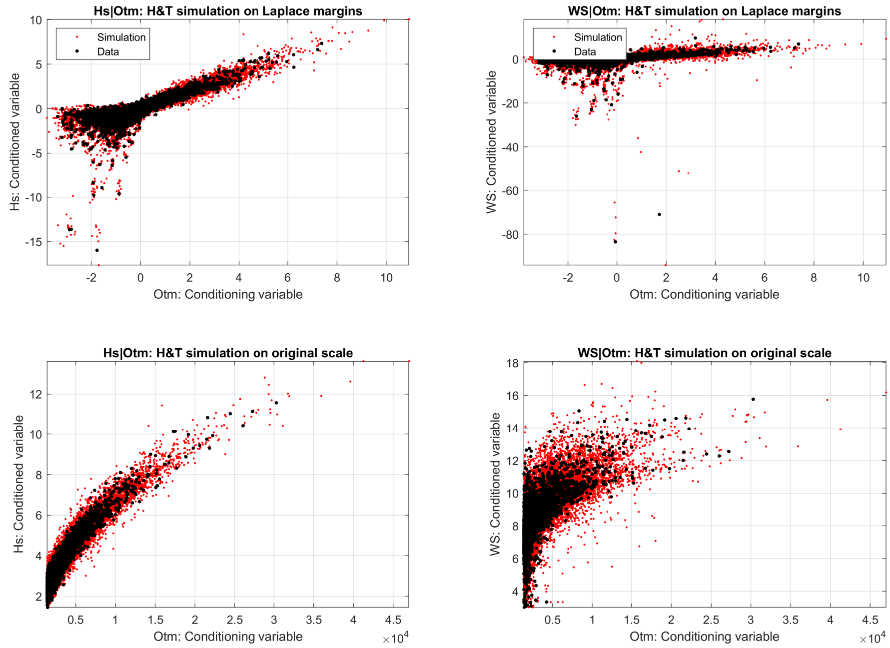

Figure 33: Comparison of original data (black) and simulation from fitted H&T model for 10× period of the data (red) on standard margins (upper plot) and original margins (bottom plot). On the original scale two different spikes can be seen at the upper right corner reflecting different marginal characteristics in T_P

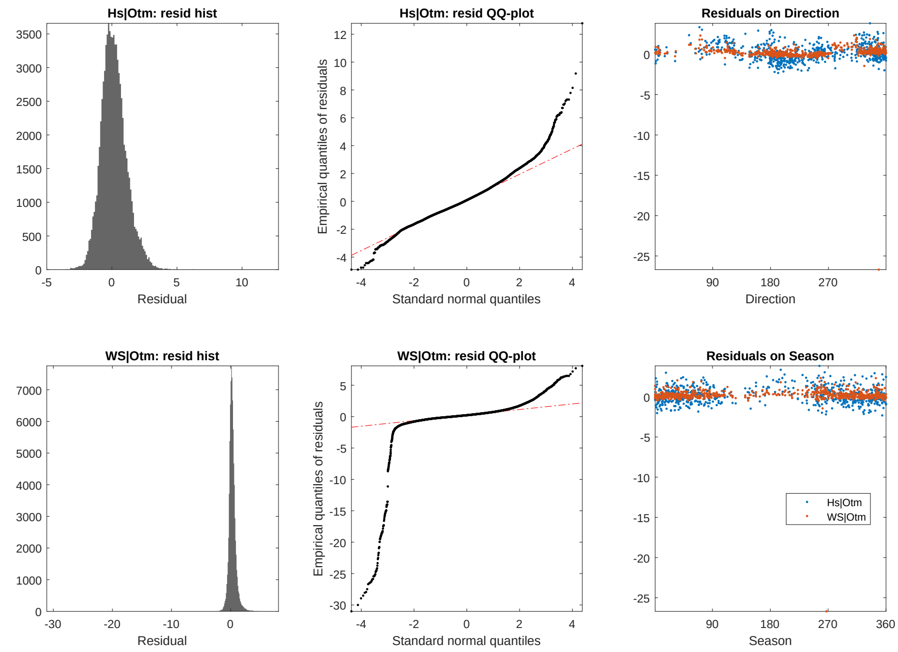

Figure 34: Diagnostic of the residuals from the HT fitting. Left column shows a histogram of the residuals for each assicoted variable. Middle column shows a normal QQ plot. The right panel shows residuals as a function of direction and season for each variable. It is typical that these residuals are quite skewed (not normal), which is why they are reused in the simulation procedure.

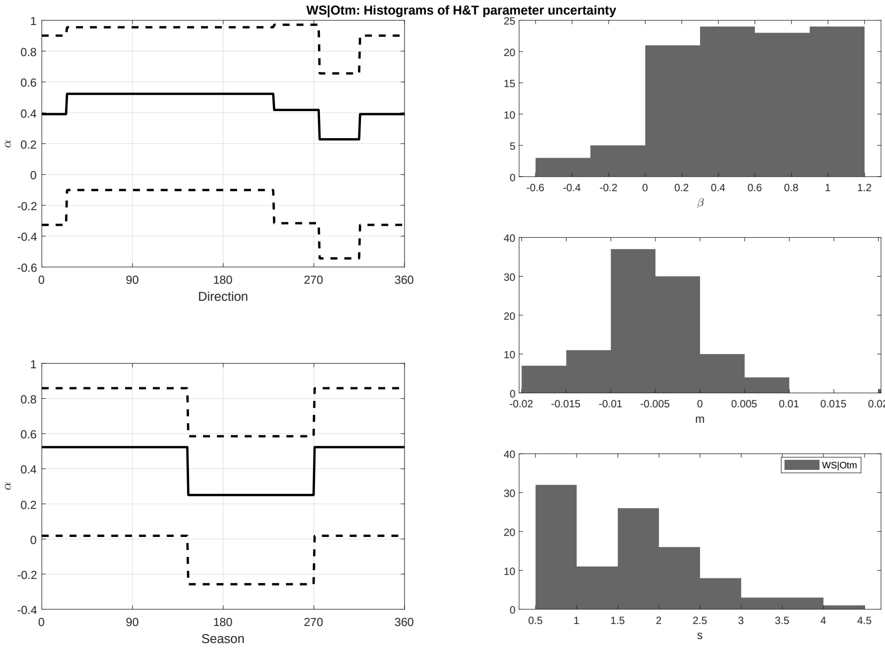

Figure 35: Histograms of the HT parameters over bootstrap re-samples for the WS conditional model in a non-stationary case. The parameter uncertainty captures, marginal (bootstrap and NEP) and conditional (bootstrap and HT NEP) uncertainty.

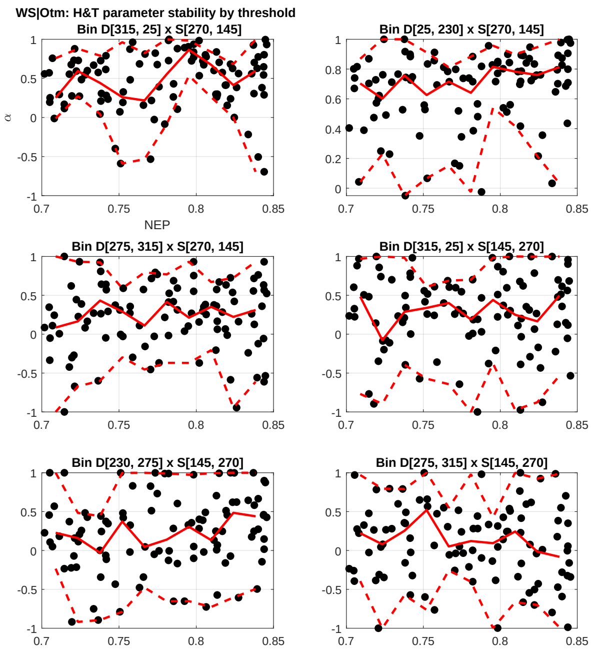

Figure 36: HT parameter $\boldsymbol{\alpha}$ as a function of the HT NEP for the WS conditional model. Black dots show individual bootstrap estimates, red lines are local binned median, 2.5 and 97.5 percentile estimates. A well behaved model should be stable over a range of NEP's

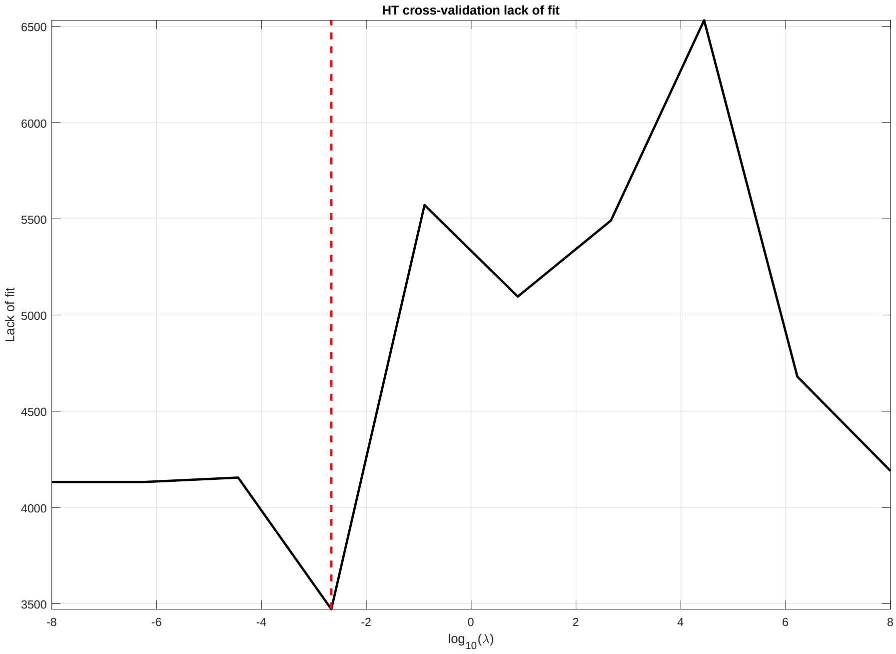

Figure 37: Cross Validation plot showing lack of fit against chosen smoothness ν of H&T parameter $\boldsymbol{\alpha}$. Low indicates good prediction performance. The red line indicates the optimal choice.

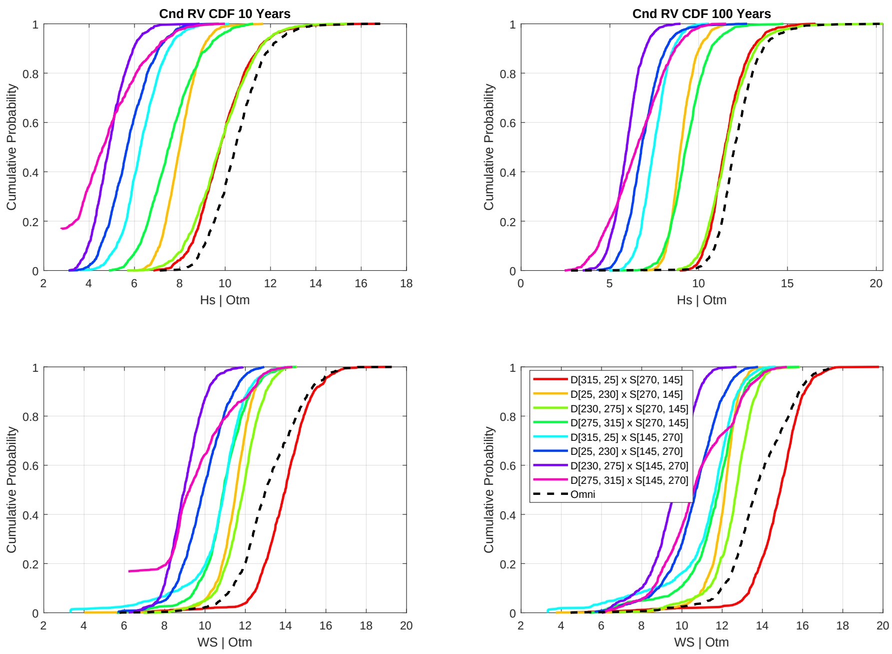

Figure 38: Conditional 10 (first column) and 100 (second year) return value CDF's of $p(H_S|\text{Otm})$ and $p(T_P|\text{Otm})$. CDFs for each covariate bin are shown using coloured lines. Black line shows the omni-directional estimate.

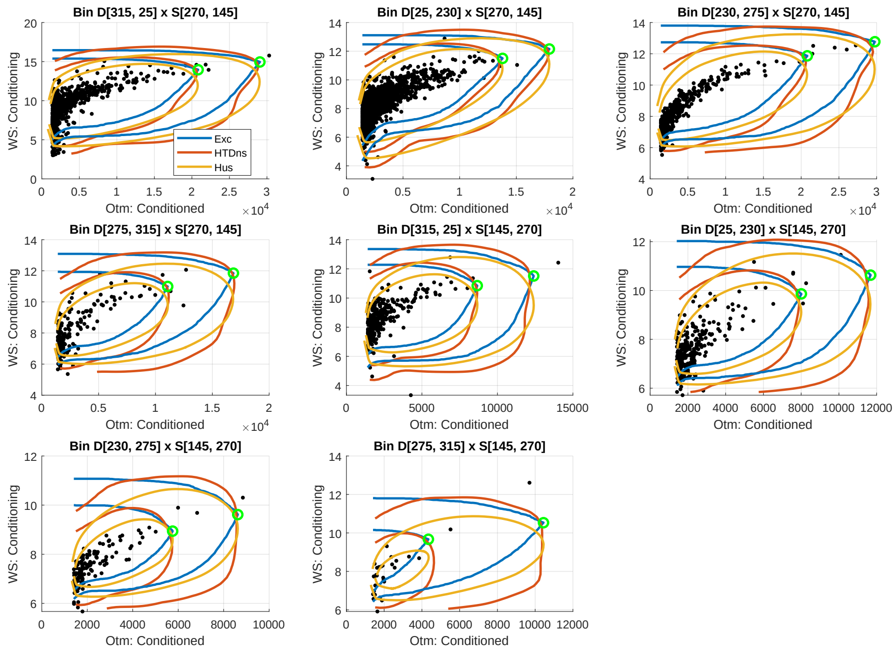

Figure 39: Comparison of contour methods for each bin sector. Contours are given for the 10 and 100 year return period. Different methods are shown using coloured lines. The green circle shows the lock point at each probability level through which all the contour methods have to pass.

A Appendix 1 : Model details

This section contains details of the underlying statistical models used.
Consider a sample $\{\dot{y}_{i1}, \dot{y}_{i2}\}_{i=1}^{n}$ of n pairs of values of peaks over threshold $\dot{Y}_1$ for a conditioning variate, and an associated conditioned variate $\dot{Y}_2$. Further, let $\{\boldsymbol{x}_i\}_{i=1}^{n}$ be the values of an associated covariate on some covariate domain $\mathcal{X}$; in this work, we assume a single directional covariate with $\mathcal{X} = [0°, 360°)$. Each pair $\dot{y}_{i1}, \dot{y}_{i2}$ is allocated to one of m covariate (directional) bins $\{C_k\}_{k=1}^{m}$ by means of an allocation vector A such that $k = A(i)$ and $\mathcal{X} = \bigcup C_k$ forms a partition. All pairs with the same covariate bin C_k are assumed to have common extreme value characteristics. The dependence structure between the variates on a transformed standard Gumbel scale is then estimated using a stationary conditional extremes model.

A.1 Marginal model

For response data $\{\dot{y}_{di}\}_{i=1}^{n}$ (where $d \in \{1, 2\}$), the likelihood function for the marginal PPC model is as follows for an interval of choices for λ_d:

$$\mathcal{L}_d(\tau_d) = \prod_{k=1}^{m} \prod_{\substack{i;A(i)=k;\\ \dot{y}_{id}>\psi_{dk}(\tau)}} \frac{1}{\nu_{dk}(\tau_d)} \left[1 + \frac{\xi_d(\tau_d)}{\nu_{dk}(\tau_d)} \left[\dot{y}_{id} - \psi_{kd}(\tau_d)\right]\right]^{-1/\xi_d(\tau_d)-1}.$$

$$\ell(\tau_d, \lambda_d) = \log \mathcal{L}(\tau_d) + \lambda_d \left(\frac{1}{m} \sum_{k=1}^{m} \nu_{dk}^2(\tau_d) - \left[\frac{1}{m} \sum_{k=1}^{m} \nu_{dk}(\tau_d)\right]^2 \right).$$

An "optimal" smoothness parameter value $\widehat{\lambda}_d$ is chosen to maximise predictive performance using a "k-fold" cross-validation procedure, and the corresponding parameter estimates (still referred to as $\xi(\tau_d), \{\nu_k(\tau_d)\}$ for brevity) used for subsequent inference. τ_d is the *marginal* non-exceedence probability for the dth response.

A.2 Transformation to Standard Scale

The marginal models are used to transform from the original scale to the standard scale (Gumbel or Laplace) using the probability integral transform.

1. First we transform data above the threshold to uniform scale using the Generalised-Pareto CDF which, with parameters ξ, ν and ψ, has cumulative distribution function:

   $$F_{GP}(\dot{y}; \xi, \nu, \psi) = (1 + \frac{\xi}{\nu}(\dot{y} - \psi))^{-1/\xi}$$

   for $\dot{y} \in (\psi, \dot{y}^+]$ where $\dot{y}^+$ is the upper end point of the distribution. Below the threshold, a gamma CDF is used:

   $$F_\Gamma(\dot{y}; \omega, \kappa, l) = \frac{1}{\Gamma(\omega)} \gamma\left(\omega, \frac{\omega(\dot{y} - l)}{\kappa}\right)$$

   where $\gamma(\cdot, \cdot)$ is the lower incomplete gamma function. Note that this an orthogonal parameterisarion taken from Cox and Reid [1987].

   That is, we transform original scale data $\dot{y}$ to uniform scale u via the following:

   $$u = \begin{cases} F_{GP}(\dot{y}) & \text{if } \dot{y} > \psi; \\ F_\Gamma(\dot{y}) & \text{if } \dot{y} \leqslant \psi. \end{cases}$$

2. Given uniformly distributed u, to transform to standard **Laplace** margins we use the inverse of the standard Laplace CDF:

   $$y = F_L^{-1}(u) = \text{sign}(0.5 - u) \times \log(2 \times \min(1 - u, u));$$

where

$$\mathbf{1}(z) = \begin{cases} 1 & \text{if } z > 0; \\ 0 & \text{if } z \leqslant 0. \end{cases}$$

For reference, this is derived from the Laplace CDF:

$$F_L(z) = \mathbf{1}(z) - 0.5 \times \text{sign}(z) \times \exp(-|z|).$$

To transform the uniform data u to standard **Gumbel** margins we use the inverse of the standard Gumbel CDF:

$$y = F_{GP}^{-1}(u) = \log(\log(-u))$$

This is derived from the Gumbel CDF:

$$F_{GP}(z) = \exp(-\exp(-z)).$$

Non stationary is captured by transforming each bin in turn with its associated parameters.

A.3 Conditional model

The Gumbel-scale sample $\{y_{1i}, y_{2i}, y_{3i}...\}$ above some threshold ϕ of the conditioning variate Y_1 is used to estimate a stationary conditional extremes model with parameters $\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\mu}$ and $\boldsymbol{\sigma}$

$$(Y_2, Y_3 \ldots |Y_1 = y_{i1}) = \boldsymbol{\alpha}_{\tilde{\tau}k} y_{i1} + y_{i1}^{\boldsymbol{\beta}_{\tilde{\tau}}} \boldsymbol{W}_{\tilde{\tau}} \text{ for } y > \phi_{\tilde{\tau}k},$$

where $\boldsymbol{W}_{\tilde{\tau}} \sim N(\boldsymbol{\mu}_{\tilde{\tau}}, \boldsymbol{\sigma}_{\tilde{\tau}}^2)$ is assumed for model estimation only. Threshold $\phi(\tilde{\tau})$ is defined as the quantile of the standard Gumbel distribution with HT non-exceedance probability $\tilde{\tau}$. Parameters are estimated to minimise the negative log-likelihood

$$\tilde{\ell}(\tilde{\tau}) = \sum_{j=1}^{D-1} \sum_{k=1}^{m} \sum_{\substack{i;k=A(i) \\ y_{ij} > \phi_{\tilde{\tau}}}} \left[\log(2\pi\sigma_{\tilde{\tau}j}^2) + \frac{1}{2\sigma_{\tilde{\tau}j}^2} \left(y_{ij} - \alpha_{\tilde{\tau}jk} y_{i1} - \mu_{\tilde{\tau}j} y_{i1}^{\beta_{\tilde{\tau}j}} \right)^2 \right]$$

for each value of $\tilde{\tau}$. This is assuming a Normal distribution in the likelihood which is almost certainly not appropriate, however the distribution of $W_{\tilde{\tau}}$ is then estimated from the sample $\{r_{\tilde{\tau}i}\}$ of residuals from the fit

$$r_{\tilde{\tau}ij} = \frac{1}{\sigma_{\tilde{\tau}j}} \left(y_{ij} - \alpha_{\tilde{\tau}jk} y_{i1} - \mu_{\tilde{\tau}j} y_{i1}^{\beta_{\tilde{\tau}j}} \right) \text{ for all } i \text{ such that } y_{i1} > \phi(\tilde{\tau}).$$

Below the threshold, $\phi(\tilde{\tau})$, data are re-sampled using an empirical CDF based on the ranks of the data.

References

D.R. Cox and N. Reid. Parameter orthogonality and approximate conditional inference. *J. Roy. Statist. Soc. Series B: Methodological*, 49:1–39, 1987.

A. F. Haselsteiner, J.-H. Ohlendorf, W. Wosniok, and K.-D. Thoben. Deriving environmental contours from highest density regions. *Coastal Eng.*, 123:42–51, 2017.

J. E. Heffernan and S. I. Resnick. Limit laws for random vectors with an extreme component. *Ann. Appl. Probab.*, 17:537–571, 2007.

J. E. Heffernan and J. A. Tawn. A conditional approach for multivariate extreme values. *J. R. Statist. Soc. B*, 66:497–546, 2004.

A Huseby, E Vanem, and B Natvig. Alternative environmental contours for structural reliability analysis. *Struct. Saf.*, 54:32–45, 2015.

P. Jonathan, J. Flynn, and K. C. Ewans. Joint modelling of wave spectral parameters for extreme sea states. *Ocean Eng.*, 37:1070–1080, 2010.

P. Jonathan, K. C. Ewans, and D. Randell. Non-stationary conditional extremes of northern North Sea storm characteristics. *Environmetrics*, 25:172–188, 2014.

C. Keef, I. Papastathopoulos, and J. A. Tawn. Estimation of the conditional distribution of a vector variable given that one of its components is large: additional constraints for the Heffernan and Tawn model. *J. Mult. Anal.*, 115:396–404, 2013.

E Ross, S Sam, D Randell, G Feld, and P Jonathan. Estimating surge in extreme North Sea storms. *Ocean Eng.*, 154:430–444, 2018.

E Ross, O Astrup, E Bitner-Gregersen, N Bunn, G Feld, B Gouldby, A Huseby, Y Liu, D Randell, E Vanem, and P Jonathan. On spatial conditional extremes for ocean storm severity. *Submitted to Ocean Engineering*, 2019.