

# Marginal and conditional extreme value models, and environmental contours, using covXtreme software

Ross Towe, Emma Ross, David Randell, Philip Jonathan

July 20, 2023

## Abstract

This report describes the covXtreme MATLAB 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 model below. The extremal dependence structure between the variates on a transformed standard scale (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.

# Contents

<b>1 Stage 1: Data preparation</b>	<b>6</b>
1.1 Stage1_PeakPicking . . . . .	6
1.2 Stage1_SimulateData . . . . .	7
<b>2 Stage 2: Choose covariate bins</b>	<b>10</b>
2.1 Running Stage2 . . . . .	10
2.2 Outputs . . . . .	11
<b>3 Stage 3: Fit marginal PPC models</b>	<b>13</b>
3.1 The penalised piecewise constant model . . . . .	13
3.2 Uncertainty quantification . . . . .	13
3.3 Running Stage3 . . . . .	13
3.4 Outputs . . . . .	14
<b>4 Stage 4: Fit conditional extremes model</b>	<b>22</b>
4.1 Conditional extremes model . . . . .	22
4.2 Running Stage4 . . . . .	23
4.3 Outputs . . . . .	23
<b>5 Stage 5: Draw contours</b>	<b>29</b>
5.1 Running Stage5 . . . . .	29
<b>6 Example: Multiple Associated Variables and Covariates</b>	<b>34</b>
<b>A Appendix: Model details</b>	<b>43</b>
A.1 Marginal model . . . . .	43
A.2 Marginal transformation to standard scale . . . . .	44
A.3 Conditional extremes model of Heffernan and Tawn (2004) . . . . .	45

# 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  $\mathbf{X}$  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 the design contours reflect these.

Non-stationarity with respect to covariates  $\mathbf{X}$  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 covXtreme 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 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. For legacy reasons we can also transform to Gumbel scale, however this does not allow for modelling negative dependence cases. 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.

In order to illustrate the usefulness of the code two case studies are presented, the first of which is modelling  $H_s$  (significant wave height) and  $T_p$  (peak period) in the presence of the covariate, wave direction. This is a standard oceanographic application and aids decision makers in determining the potential loads on an offshore structure. This is the application presented throughout Sections 1 to 5. A more complicated example is presented in Section 6, whereby we examine the joint relationship of  $H_s$  and  $T_p$  as well as the OTM (overturning moment) for different wave directions and times of the year. Overturning moment is the movement of a vessel whereby it can no longer be considered to be stable.

This software was originally 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”

(ECSADES), 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].

## List Of Symbols

- $D$  = total number of dimensions / responses  $Y$ , indexed by  $d$
- $N$  = number of observations, i.e. length of  $Y$
- $i$  = index on response or covariate observation ( $\in \{1, \dots, N\}$ )
- $C$  = total number of covariates, indexed by  $c$
- $B$  = total number of covariate bins, indexed by  $b$
- $\dot{Y}$  = response data on original scale
- $Y$  = random variable representing response data on standard Laplace scale
- $X$  = random variable representing covariate data
- $R$  = number of different return-periods in analysis (return-value CDFs and contours)
- $t$  = storm-picking non-exceedance probability (Stage 1)
- $\tau$  = non-exceedance probability for marginal modelling (Stage 3)
- $\psi$  = threshold associated with non-exceedance probability  $\tau$  for marginal modelling (Stage 3)
- $\xi, \nu$  = marginal model parameters for the generalised Pareto distribution above the threshold (Stage 3)
- $\omega, \kappa, l$  = marginal model parameters for the Gamma distribution below the threshold (Stage 3)
- $\lambda$  = marginal model roughness penalty on extent to which  $\nu$  can vary by covariate bin (Stage 3)
- $\tilde{\tau}$  = non-exceedance probability for Heffernan & Tawn modelling (Stage 4)
- $\phi$  = threshold associated with non-exceedance probability  $\tilde{\tau}$  for Heffernan & Tawn modelling (Stage 4)
- $\alpha, \beta, \mu, \sigma$  = Heffernan & Tawn model parameters (Stage 4)
- $\tilde{\lambda}$  = Heffernan & Tawn model roughness penalty on extent to which  $\alpha$  can vary by covariate bin (Stage 4)

# 1 Stage 1: Data preparation

In Stage 1 we prepare  $D$ -dimensional peaks over threshold response data  $\{\hat{y}_{1i}, \hat{y}_{2i}, \dots, \hat{y}_{Di}\}_{i=1}^N$  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, both producing a data file called `Data.mat`, the format of which is described below. When data is available, for example time series observations for Hs and Tp, `Stage1_PeakPicking` should be run to identify storm peak data from the sea-state observations. 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.

**Run Scripts:** `Stage1_PeakPicking.m` OR `Stage1_SimulateData.m` (in the `Case1` folder)

**Output data:** `Output\Data.mat` with the following contents

- `Dat.Y` [ $N \times D$ ] matrix of response data, with the main (conditioned) response in the first column; and associated response data in the subsequent columns.
- `Dat.X` [ $N \times C$ ] matrix of covariate data, with each column representing a different covariate.
- `Dat.RspLbl` [ $1 \times D$ ] cell array containing string descriptions of the responses, e.g. `{'Hs', 'Tp'}`.
- `Dat.CvrLbl` [ $1 \times C$ ] cell array containing string descriptions of the covariates.
- `Dat.IsPrd` [ $1 \times C$ ] boolean vector marking periodicity of covariates (0 = non-periodic; 1 = periodic).

You can also skip Stage 1 entirely by manually populating an equivalent `Data.mat` data file with a format identical to that described above.

Figures generated when running the covXtreme software are stored in a `Figures` subdirectory, with the stage they are generated in indicated by prefix `StgX_` in the file name.

## 1.1 Stage1\_PeakPicking

This script converts time series data into peaks-over-threshold data, suitable for modelling with the generalised Pareto distribution.

An example data file called `CNS_ResponseData.mat` is provided with this software. If you wish to use this data file, its location should be provided to the `load()` command at the start of the MATLAB script. If you are instead using a dataset of your own, not in `.mat` format, you should replace this line with a call to e.g. the `readcsv()` function so that you can enter your data properties to the relevant inputs described below.

**Inputs:**

- `RspLbl` [ $D \times 1$ ] 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 defined on [0,1] which is used to define the threshold for storm-identification.

Suppose that we have set the main response  $Rsp=Hs$  and associated response  $Asc=Tp$ ; 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  $Hs$ ) and take *associated* observations as peaks over threshold for  $T_P$ .

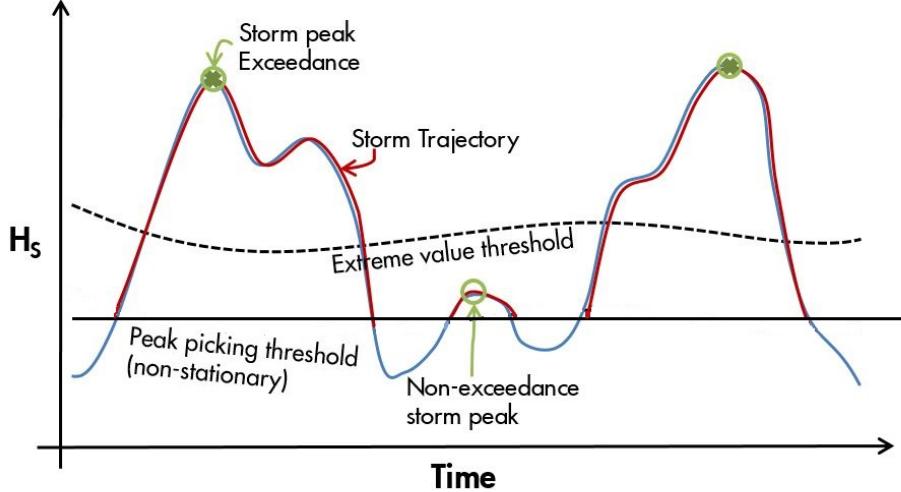


Figure 1: Peak Picking Illustration

An example of storm peak-picked data is shown for North Sea data in Figures 2 and 3. A quantile level of  $t = 0.6$  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. `Mdl.nDmn` [scalar] the number of response variables you want to simulate
2. `Mdl.nDat` [scalar] the number of observations you want to simulate
3. `Mdl.nBin` [scalar] the number of covariate bins you want (common to both margins if `nDmn > 1`)
4. `Mdl.DrcEdg` [ $1 \times nBin$ ] vector of edges of covariate bins on  $[0, 360]$  (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. `Mdl.MM.Shp` [ $1 \times nBin$ ] vector of GP shape parameters for each covariate bin
2. `Mdl.MM.Scl` [ $1 \times nBin$ ] vector of GP scale parameters for each covariate bin
3. `Mdl.MM.Thr` [ $1 \times nBin$ ] vector of GP thresholds for each covariate bin
4. `Mdl.Rat` [ $1 \times nBin$ ] vector of Poisson rate parameters for each covariate bin

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. `Mdl.Jnt.Mth` [ $1 \times nBin$ ]: Choice of dependence model: multivariate normal `MVN`; logistic `LGS`; or asymmetric logistic `ASL`

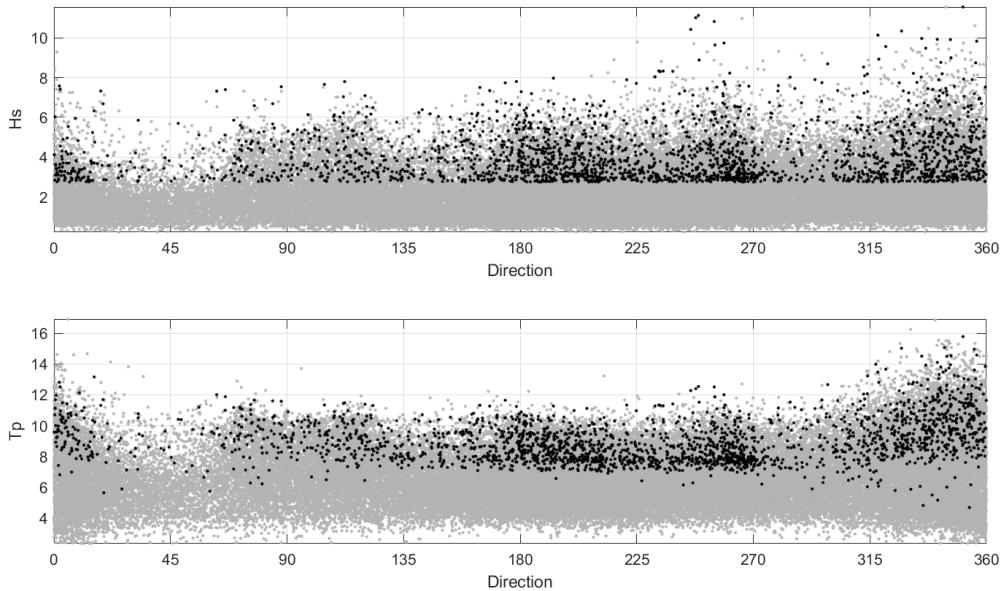


Figure 2: Marginal Hs and Tp as a function of direction for North Sea data. Storm peaks shown in black, all sea states in grey.

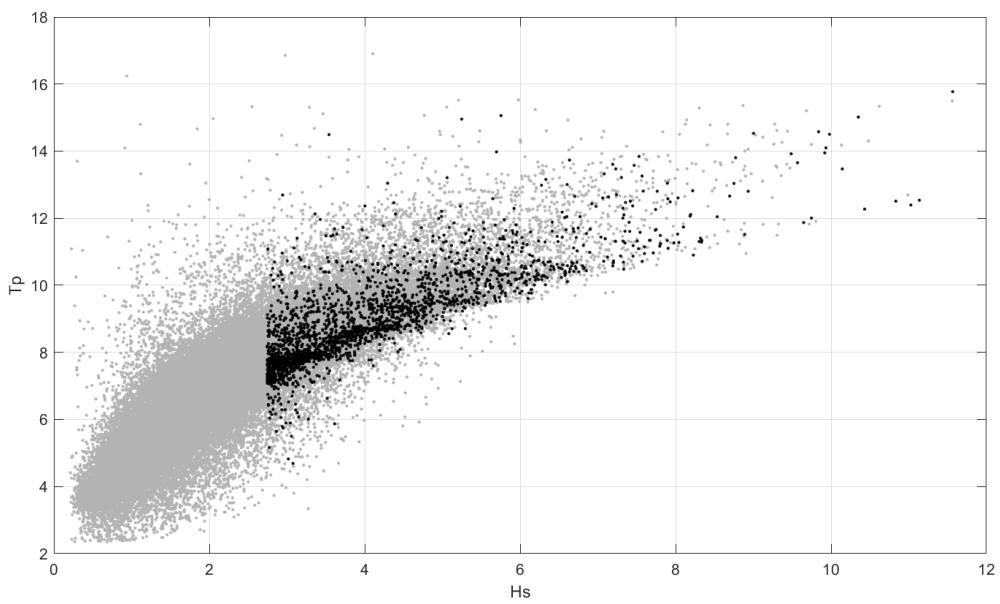


Figure 3: Joint distribution Hs and Tp for North Sea data. Storm peaks shown in black, all sea states in grey.

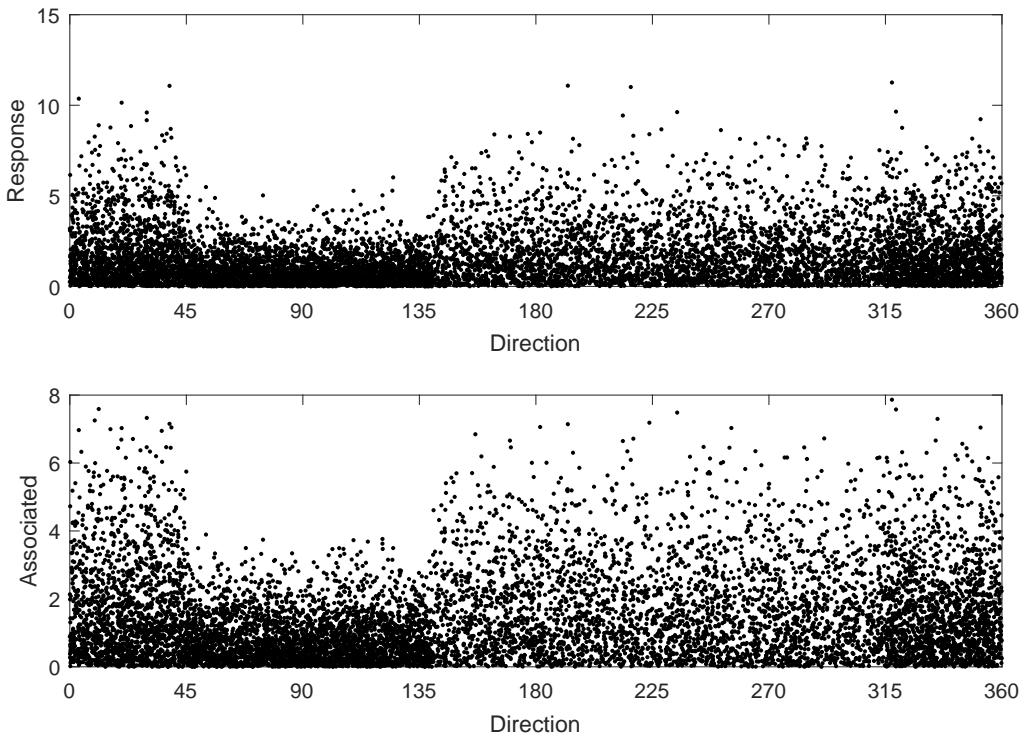


Figure 4: Example simulated bivariate data that is a function of a covariate.

## 2. Associated parameters:

- MVN : dependence parameter  $\text{Rho} \in [0, 1]$  per covariate bin  
 $\text{Md1.Jnt.Rho}$  [ $1 \times \text{nBin}$ ]
- LGS : dependence parameter  $\text{Alp} \in [0, 1]$  per covariate bin  
 $\text{Md1.Jnt.Alp}$  [ $1 \times \text{nBin}$ ]
- ASL : dependence parameter  $\text{Alp}$  as above and weighting parameters  $\text{Theta}$  (one for each response/margin)  $\in [0, 1]$  per covariate bin setting the proportion of ‘random’ points off of the logistic dependence  
 $\text{Md1.Jnt.Alp}$  [ $1 \times \text{nBin}$ ]  
 $\text{Md1.Jnt.Theta}$  [ $1 \times \text{nBin}$ ]

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 2.

## 2 Stage 2: Choose covariate bins

At the start of Stage 2, peak-picked data  $\{\hat{y}_{di}\}_{i=1}^N$  for each marginal response  $d \in \{1, \dots, D\}$  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`

**Inputs:** `BinEdg` [ $1 \times C$ ] cell array storing bin edges for each covariate.

In the case of a single covariate, bin edges should be provided to input `BinEdg` in  $\{[\dots]'\}$  format. Note that we need to transpose (' operation in MATLAB) to put the bin edges into long vector format. In the case of multiple covariates; bin edges should be provided in  $\{[\dots]', [\dots]'\}$  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 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  $\{[\min(\text{Cvr}), \max(\text{Cvr})]'\}$ .

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.

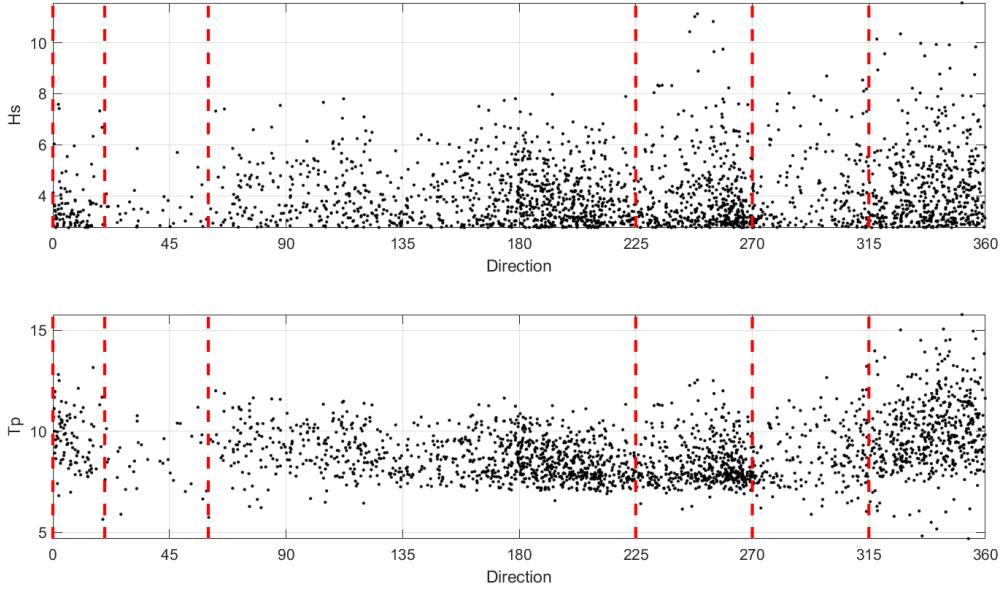


Figure 5: Example bin allocation. Bins are chosen at [0,25,60,225,270,315] degrees. Storm peak data shown in black, chosen bin edges are shown with red dashed lines.

## 2.2 Outputs

Running Stage 2 creates a MATLAB data file (.mat extension) called **Data**, stored in a folder called **Output**.

The following Figures are also generated and saved as .pngs in the **Figures** folder. Figure 5 illustrates the user's bin choice as red lines on top of scatter plots of response(s) against covariate(s). Then for each associated variable, a figure like Figure 6 is produced, containing scatter plots of the associated response on the  $y$  axis and main/conditioned response on the  $x$  axis, broken out by covariate bin. The dependence structure illustrated by these subplots (non-stationary with respect to covariates) is what we aim to model via marginal and conditional extremes modelling in subsequent Stages 3 and 4.

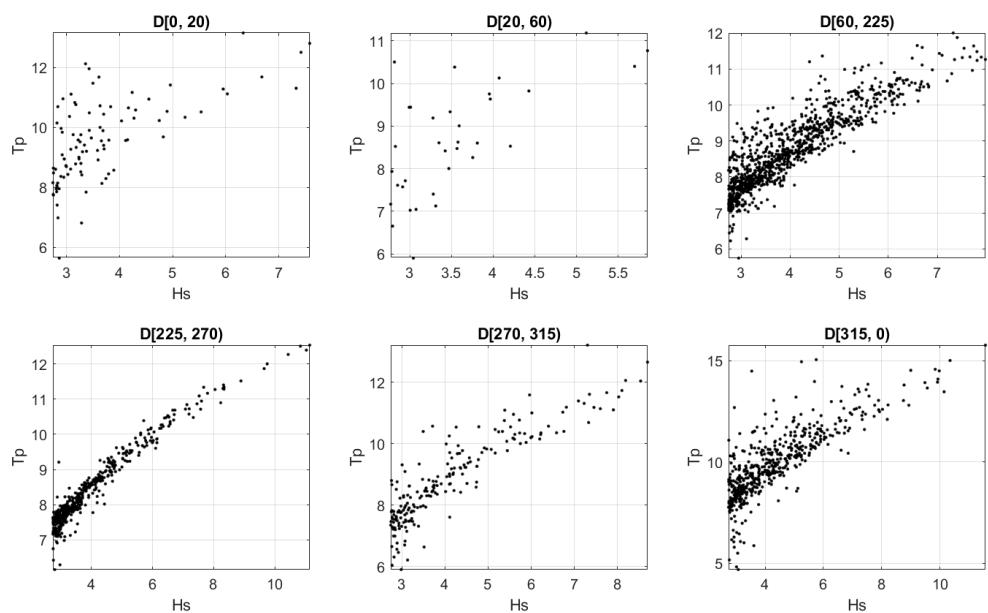


Figure 6: Scatter plots of storm peak responses of  $H_s$  plotted against associated values of  $T_p$  broken out by bin.

### 3 Stage 3: Fit marginal PPC models

#### 3.1 The penalised piecewise constant model

Stage 3 fits a PPC (penalised piecewise constant) 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  $b$ , the extreme value threshold  $\psi_b(\tau)$  is assumed to be a quantile of the Gamma distribution fitted to all data in that bin, with specified non-exceedance probability  $\tau$  that is constant across bins. Threshold  $\psi_b(\tau)$  is estimated with no smoothing across bins.

Threshold exceedances are assumed to follow the GP distribution with shape  $\xi(\tau)$  and scale  $\nu_b(\tau)$ . Since estimation of the shape parameter is particularly problematic, the shape parameter is assumed constant (but unknown) across covariate bins. The shape parameter of the GP is particularly difficult to estimate and many applications is assumed to be constant to limit unrealistic extrapolations from the model. The extent to which the GP scale varies across bins is controlled by smoothness parameter  $\lambda$ . Then parameters  $\{\xi(\tau), \{\nu_b(\tau)\}\}$  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  $l_b(\tau)$ , shape  $\omega_b(\tau)$ , and scale  $\kappa_b(\tau)$ , 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 smoothness of the GP scale parameter 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 Stage 3 script, we recommend choosing Laplace) 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 are NEP, CV.SmthLB and CV.SmthLB.

- **iDmn** [scalar] specify the response upon which to fit marginal model
- **NEP** [ $1 \times 2$ ] non-exceedance probability range, should be in  $(0, 1)$
- **nBoot** [scalar] number of bootstrap re-samples - must use the 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 ( $\log_{10}$ ) for smoothness range
- `CV.SmthUB` [scalar] upper bound ( $\log_{10}$ ) for smoothness range
- `MarginType` [string] specify the standard margin scale on which the Heffernan & Tawn model will be fitted (options are ‘Laplace’ or ‘Gumbel’, with the default set as Laplace)

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: `Mrg should be an nDmn x 1 Marginal Model.`

Note that the input settings for each margin can generally differ, however the number of bootstraps `nBoot` and standard margin `MarginType` must be common to all scripts, for consistency when fitting the conditional model in Stage 4.

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 detail on this process is given as part of the figure).

Details which should be considered carefully, to ensure a quality marginal fit, are contained in boxes.

### 3.4 Outputs

Each application of the marginal model-fitting procedure (Stage 3) to variate/dimension  $d$  creates a MATLAB data file (.mat extension) called `MM#` (with dimension  $d$  in the place of #) and stored in a folder called `Output`. Before you move onto running Stage 4, you should verify that you have one such file for each dimension. The file contains the following data:

- `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.nBoot` [scalar] number of bootstraps used
- `MM.RtrPrd` [ $1 \times R$ ] return periods
- `MM.Bn` covariate bin structure (created in Stage 2)
- `MM.Sc1` [ $B \times nBoot$ ] Generalised Pareto scale parameter
- `MM.Shp` [ $nBoot \times 1$ ] Generalised Pareto shape parameter
- `MM.Omg` [ $B \times nBoot$ ] Gamma shape parameter
- `MM.Kpp` [ $B \times nBoot$ ] Gamma scale parameter
- `MM.GmmLct` [ $B \times 1$ ] Gamma location parameter
- `MM.NEP` [ $B \times nBoot$ ] non exceedance probability
- `MM.Thr` [ $B \times nBoot$ ] exceedance threshold
- `MM.Rat` [ $B \times nBoot$ ] Rate of occurrence

- **MM.BSInd** [ $N \times n_{\text{Boot}}$ ] 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.MarginType** [string] distribution used to transform to standard margins
- **MM.RVSml.nRls** [scalar] fixed number of MC realisations, set at 10,000
- **MM.RVSml.I** [ $B \times \text{MM.RVSml.nRls} \times \text{MM.nRtr}$ ] indicator matrix to decide which bootstrap to sample from
- **MM.RVSml.A** [ $B \times \text{MM.RVSml.nRls} \times \text{MM.nRtr}$ ] matrix containing the bin allocation for each of the bootstraps
- **MM.RVSml.Unf** [ $B \times \text{MM.RVSml.nRls} \times \text{MM.nRtr}$ ] simulated return values for responses on probability scale (uniform margins)
- **MM.RVSml.Org** [ $B \times \text{MM.RVSml.nRls} \times \text{MM.nRtr}$ ] simulated return values for responses on original scale

simulated return values for responses on uniform scale (Uniform margins).

The following figures illustrate the result of PPC model fitting for the North Sea example on the  $H_s$  margin.

The leftmost panel of Figure 7 shows the original response data plotted against covariate  $\mathbf{X}$ . The blue lines represent a 95% bootstrapped uncertainty interval on the location of the threshold and incorporate two sources of randomness originating from bootstrap re-sampling and from drawing non-exceedance probability  $\tau$  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  $\tau$  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 Laplace margins (the process followed using the PIT).

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).

Figure 8 gives 95% bootstrap uncertainty intervals on the non-stationary GP scale as a function of covariate as well for the non-stationary shape and scale parameters of the Gamma distribution. Figure 9 gives a histogram of (stationary) GP shape parameter estimate. Again, these incorporate both bootstrap resampling uncertainty and NEP sampling uncertainty. 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.

Figure 10 illustrates the cross-validation on roughness penalty  $\lambda$ , via a lack-of-fit plot for values within range  $[\text{CV.SmthLB}, \text{CV.SmthUB}]$ .

If the red line, indicating the optimal choice of  $\lambda$  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`.

The quality of model fit within each covariate bin can be assessed using Figure 11. Red dots outside the bootstrap uncertainty 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.

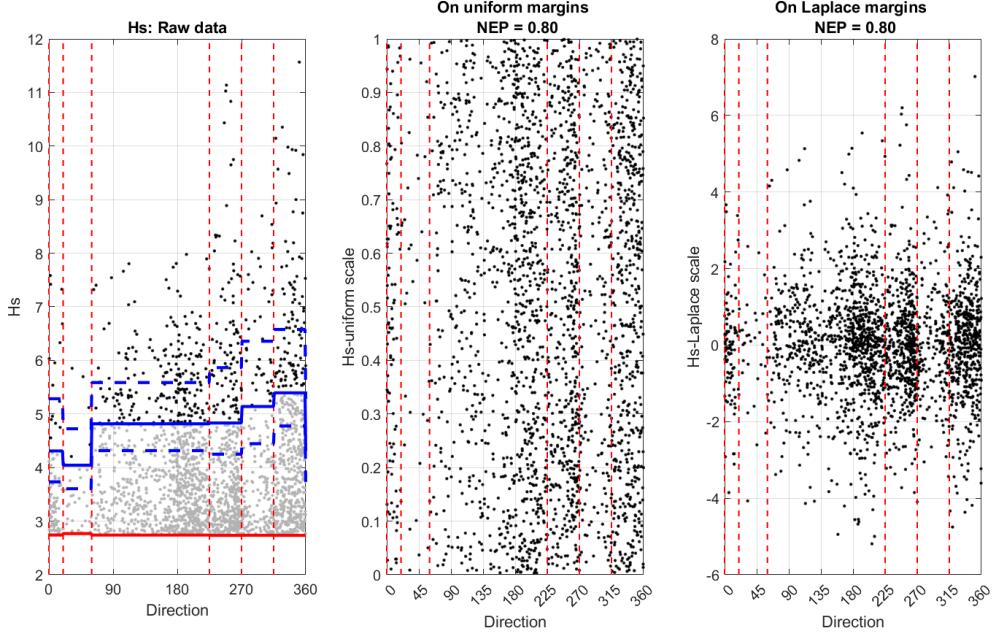


Figure 7: Left panel shows sea-state data in grey and storm peaks in black. Bin edges are indicated by dashed red lines. 2.5, 50 and 97.5 percentiles of estimated threshold across all bootstraps are plotted with blue lines. The Gamma location parameter is plotted in solid red. Middle and right panels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

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 top panel a reasonable choice might be [0.55,0.75], in the bottom panel a narrower range might be chosen, say [0.55, 0.65]. The right limit can usually be chosen as the last point before which the bootstrap uncertainty 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 maximum of 10 and 100 years cumulative distribution functions for each covariate (here, directional) sector as well as the omni-directional distribution. 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.

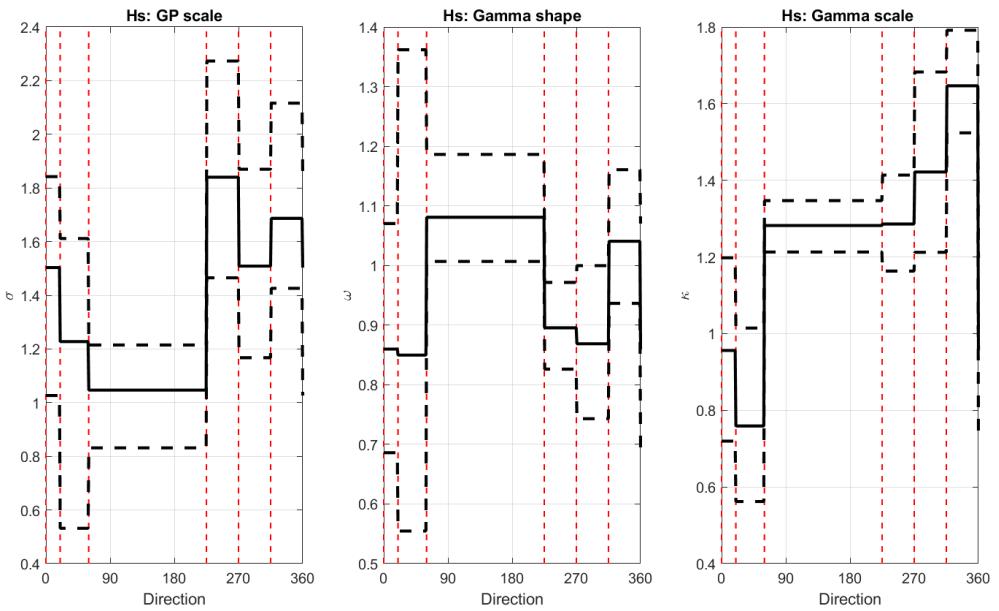


Figure 8: Black lines show 2.5, 50 and 97.5 percentiles of GP scale and the Gamma shape and scale as a function of direction respectively. Red lines show bin edges.

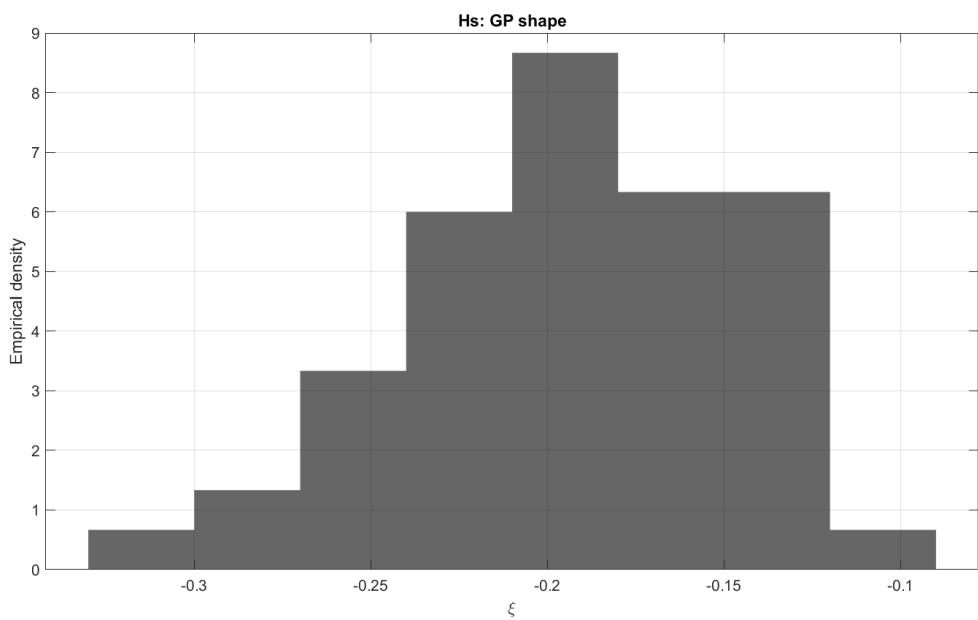


Figure 9: Histogram of estimated GP shape parameter, assumed constant with respect to the covariate in this case direction.

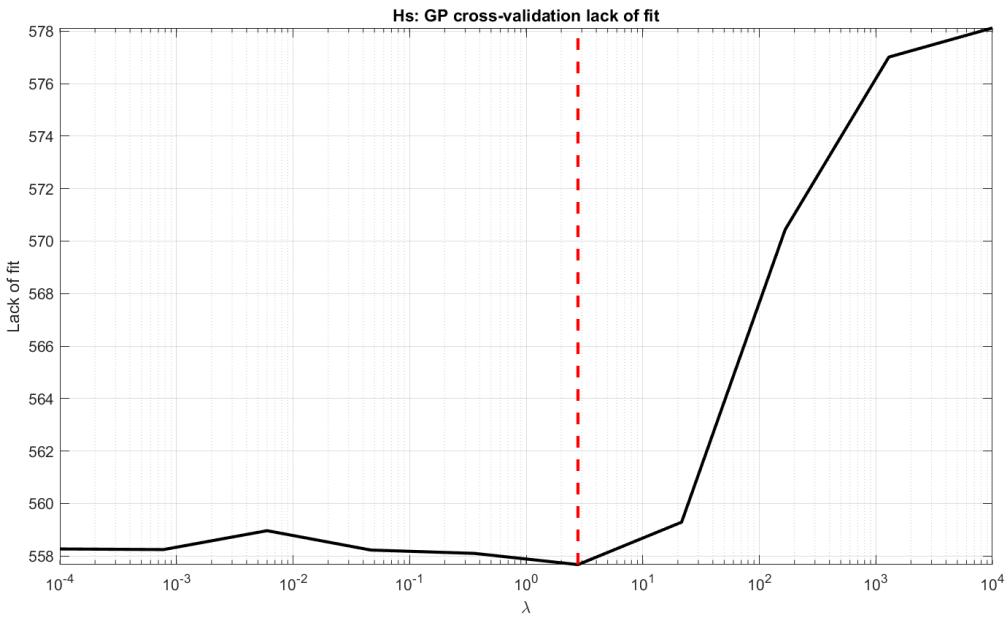


Figure 10: Cross Validation plot showing lack of fit against chosen smoothness  $\lambda$  of GP scale. Low indicates good prediction performance. The red dashed line indicates the optimal choice.

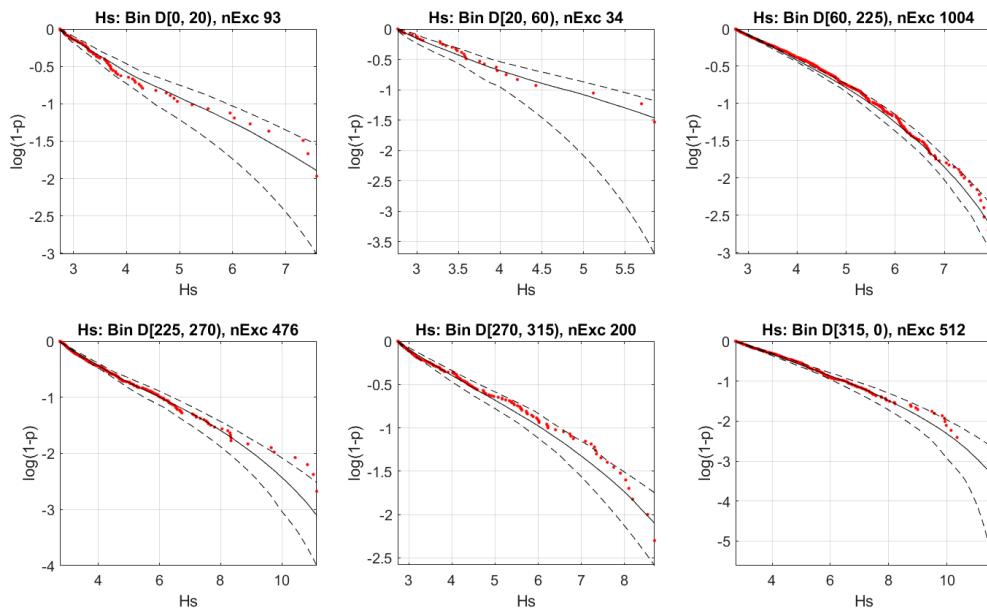


Figure 11: Diagnostic for quality of model-fit, broken out by covariate (here, directional) sector. Red dots show storm peaks, black lines are 2.5, 50 and 97.5 percentiles of model prediction over bootstraps. Red dots inside the bootstrap uncertainty limits of the model indicate a good fit.

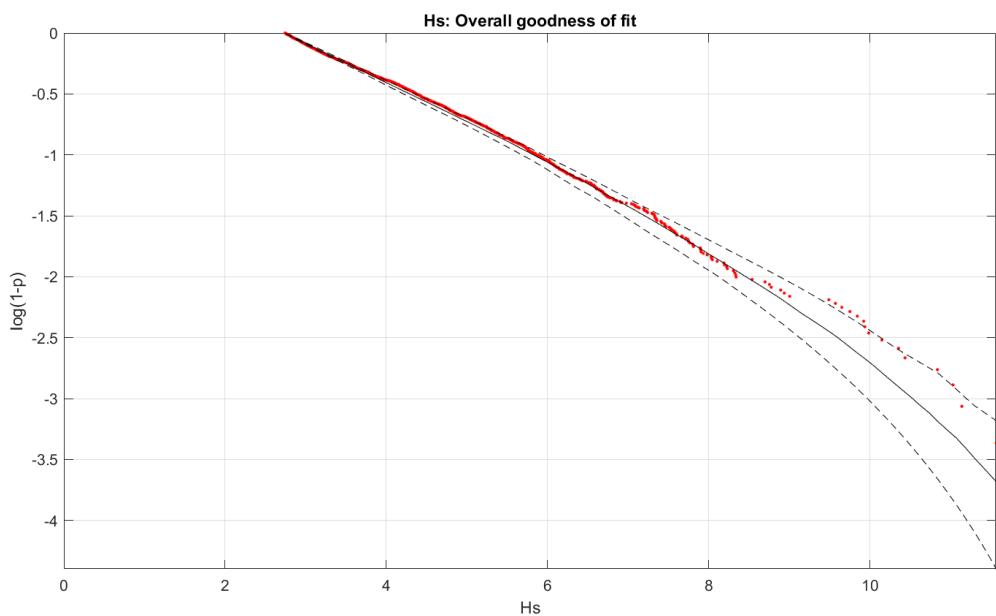


Figure 12: Diagnostic for overall quality of model-fit. Red dots show storm peaks, black lines are 2.5, 50 and 97.5 percentiles of model prediction over bootstraps. Red dots inside the bootstrap uncertainty limits of the model indicate a good fit.

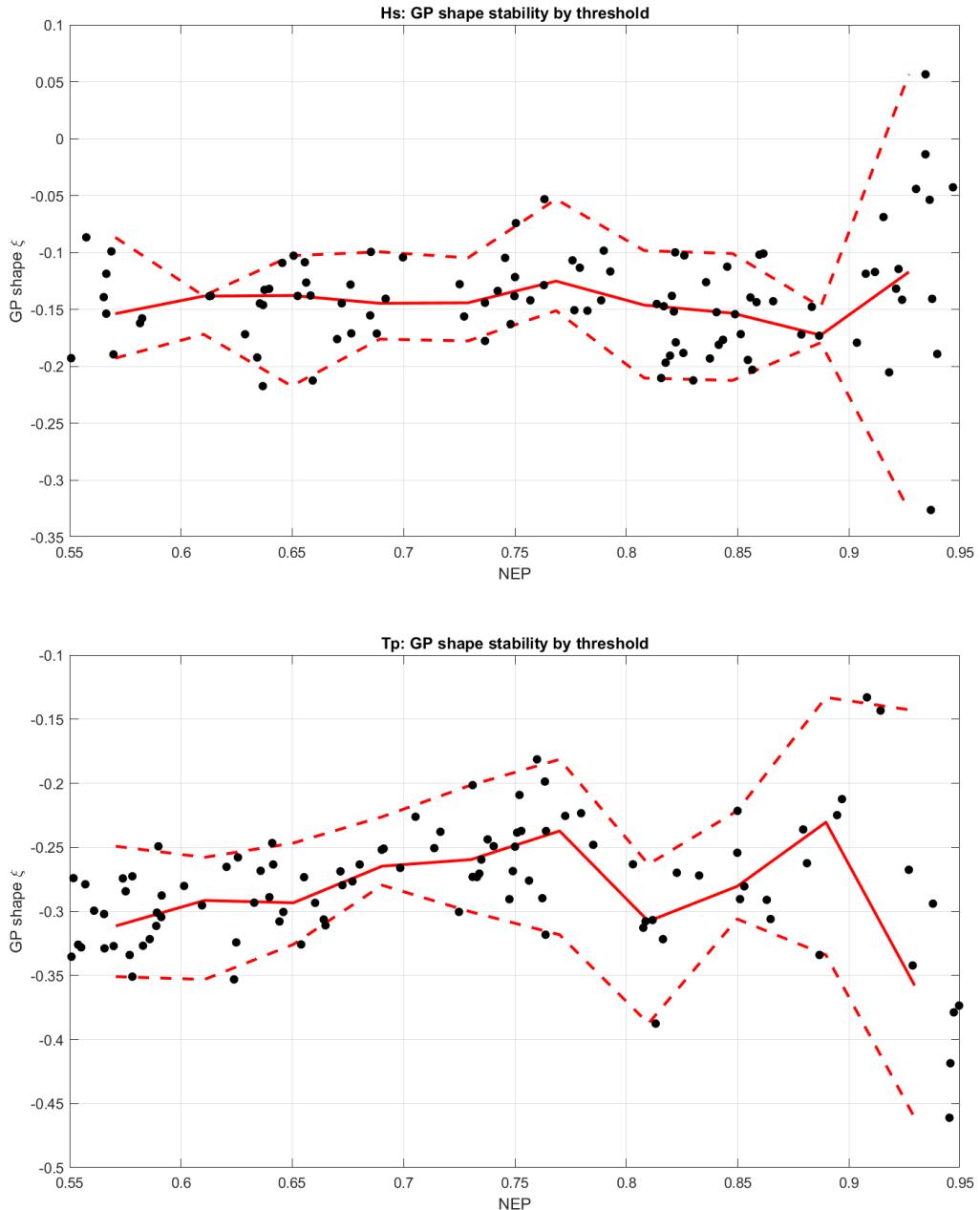


Figure 13: GP shape  $\xi$  as a function of the NEP for 2 responses Hs (top plot) and Tp (bottom plot) from the North Sea data. Black dots show individual bootstrap estimates, red lines are local binned median, 2.5 and 97.5 percentile estimates. A well behaved model should be stable over a range of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

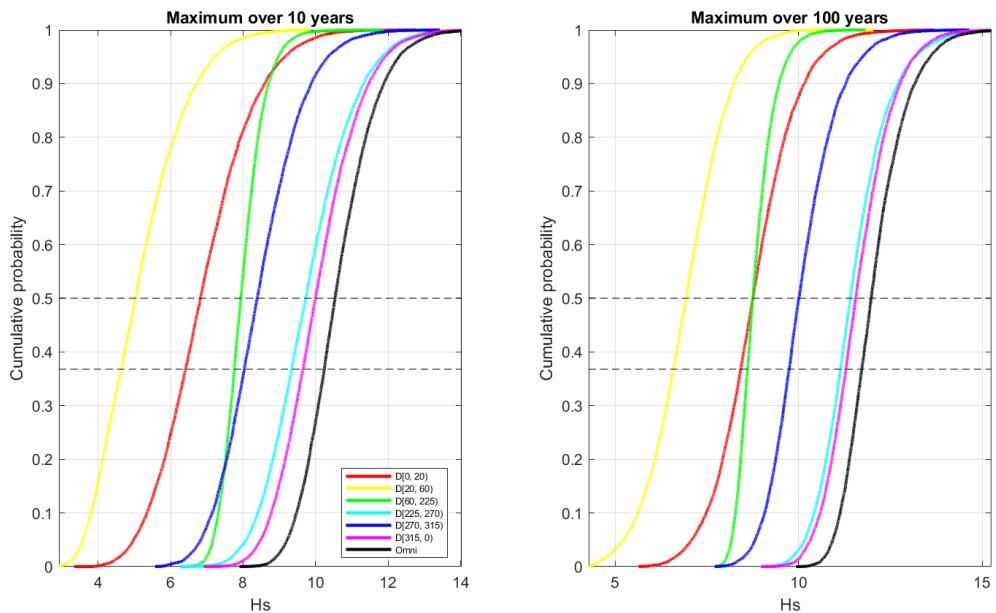


Figure 14: Marginal 10 year (upper plot) and 100 year (lower plot) return value CDFs (Hs). Directional sectors are shown using coloured lines. The black line shows the omni-directional estimate. The two horizontal dashed black reference lines at 0.368 and 0.5 that correspond to the most probable quantile ( $\exp(-1)$ ) and the median quantile respectively.

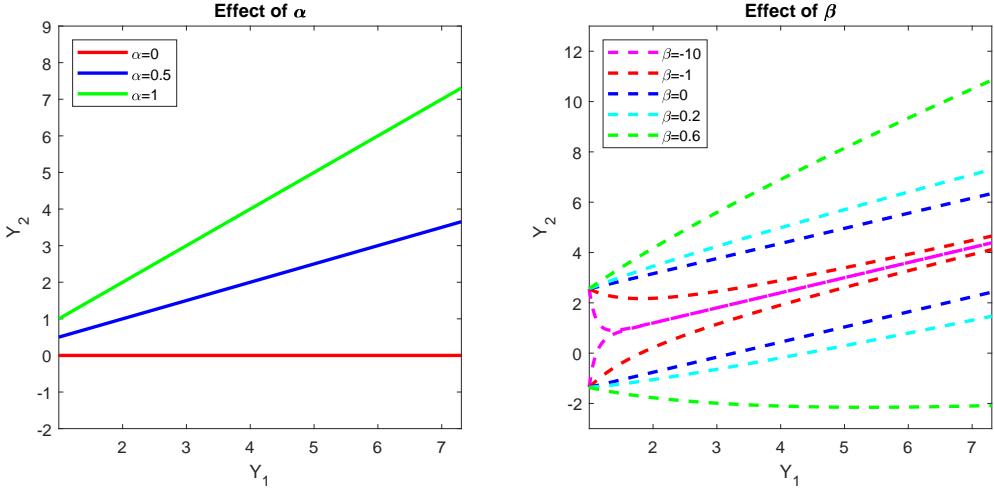


Figure 15: Illustration of the impact of H&T parameters  $\alpha$  and  $\beta$  on the structure of dependence between two random variables ( $Y_1$ ,  $Y_2$ ) on standard margins.

## 4 Stage 4: Fit conditional extremes model

### 4.1 Conditional extremes model

For simplicity a bivariate version of the conditional extremes model is presented below with a  $D$ -dimensional version presented in the appendix.

Consider a pair of random variables  $Y_1$  and  $Y_2$  on standard scale in this case Laplace margins. We take  $Y_1$  to be the conditioning margin of the conditional extremes model through assuming that

$$Y_2|(Y_1 = y_1) = \alpha_{\tilde{\tau}b}y_1 + y_1^{\beta_{\tilde{\tau}b}}W_{\tilde{\tau}b} \text{ for } y_1 > \phi_{\tilde{\tau}b}. \text{eq : HTFramework}$$

where the conditioning variable  $Y_1$  is above the threshold  $\phi_{\tilde{\tau}b}$ , which is defined as the quantile of the standard Laplace distribution with non-exceedance probability  $\tilde{\tau}$  for covariate bin  $b$ . The conditional extremes model has two parameters  $\alpha_{\tilde{\tau}b} \in [-1, 1]$  and  $\beta_{\tilde{\tau}b} \in (-\infty, 1]$ . The parameter  $\alpha_{\tilde{\tau}b}$  captures the extent of positive linear dependence between the associated and conditioned variable (with a stronger positive relationship as  $\alpha_{\tilde{\tau}b} \rightarrow 1$ ). We also allow the user change the limits of  $\alpha_{\tilde{\tau}b}$  to be greater than 1. This follows recent advancements in the literature providing improved estimation for asymptotic dependence cases at subasymptotic levels.

The  $\beta_{\tilde{\tau}b}$  parameter of the conditional extremes model captures the spread of data around that linear relationship, with large negative values indicating a very tight distribution of data particularly for higher values of the conditioning variable; and positive values indicating a large degree of variance around the relationship. An illustration of this varying nature of dependence which the conditional extremes model can capture is given in Figure 15.

Finally,  $W_{\tilde{\tau}b}$  represents the residual distribution of the conditional extremes model. For inference we take the false working assumption that  $W_{\tilde{\tau}b}$  follows a generalised Gaussian (GG) distribution. The GG distribution has three parameters  $(\mu_{\tilde{\tau}b}, \sigma_{\tilde{\tau}b}, \delta)$ , where  $\mu_{\tilde{\tau}b} \in (\infty, \infty)$ ,  $\sigma_{\tilde{\tau}b} \in [0, \infty)$  and  $\delta \in [0, \infty)$ . The GG distribution has an extra parameter  $\delta$  compared to the Gaussian distribution. The  $\delta$  acts as a shape parameter and allows the residual distribution to follow a number of standard distributions, for example if  $\delta = 1(2)$  the residuals follow a Laplace(Gaussian) distribution respectively.

Model-fitting therefore corresponds to estimating  $\{\alpha_{\tilde{\tau}b}, \beta_{\tilde{\tau}b}, \mu_{\tilde{\tau}b}, \sigma_{\tilde{\tau}b}\}$  given a sample of values for  $Y_1, Y_2$  such that  $Y_1 > \phi_{\tilde{\tau}b}$ .

All of  $\phi_{\tilde{\tau}b}, \alpha_{\tilde{\tau}b}, \beta_{\tilde{\tau}b}, \mu_{\tilde{\tau}b}$  and  $\sigma_{\tilde{\tau}b}$  are in principle functions of covariates (indexed by  $b$ ). We recommend incorporating covariates into  $\alpha_{\tilde{\tau}b}$  and  $\mu_{\tilde{\tau}b}$  before incorporating non-stationarity into the other parameters

of the conditional extremes model. Using the conditional extremes model, we simulate joint extremes on the standard Laplace scale, and transform these realisations to the original scale using the probability integral transform once more. As part of this simulation we drop the false working assumption that the residuals  $W_{\tilde{\tau}b}$  follow a generalised Gaussian distribution and instead empirically resample from the residuals.

## 4.2 Running Stage4

Running Stage4 fits a Heffernan and Tawn conditional extreme value model. Marginal model data and parameters (`Output\MM1.mat` and `Output\MM2.mat`) are loaded from Stage 3.

**Run Scripts:** `Stage4_FitHeffernanTawn.m`

**Output files:** `Output\HT.mat`

**Inputs:**

- **HTNEP** [scalar] conditional non exceedence probability range, when using the Gumbel margin type make sure that verb—`HTNEP`— >  $\exp(-\exp(-0)) = 0.368$  or the marginal transformation will fail;
- **NonStationary** [ $4 \times 1$ ] If 0: use a stationary for the equivalent parameter in H&T model; if 1: fit penalised-piecewise-constant (non-stationary) using the same bins as the marginal analysis;
- **CV.CVMth** [boolean] If 0: Only cross validate roughness penalty for original dataset (fast); or 1: Cross Validate smoothness for every bootstrap re-sample (slow);
- **CV.nCV** [scalar] number of cross validation groups;
- **CV.nSmth** [scalar] number of roughnesses tried in CV;
- **CV.SmthLB** [scalar] lower bound ( $\log_{10}$ ) for set of candidate smoothness penalties;
- **CV.SmthUB** [scalar] upper bound ( $\log_{10}$ ) or set of candidate smoothness penalties;
- **SampleLocalResid** [boolean] If this is set to true (or 1), when simulating under H&T model residuals are resampled locally (from the current covariate bin); if false (or 0), residuals are sampled globally, i.e. from any bin.

Note that sampling residuals locally (setting `SampleLocalResid` = true) effectively results in the residual part of the H&T model being non-stationary, thus improving the fit/simulation procedure. That said, in the presence of bins with very few observations, we advise 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 resamples is inherited from the marginal model settings. For this stage to run successfully, you must have used the same number of bootstraps in each marginal model run (already highlighted in Section 3.3).

## 4.3 Outputs

Running Stage 4 produces a MATLAB data file called `HT.mat` with the following contents:

- **HT.Alp** [ $B \times (D - 1) \times \text{nBoot}$ ]  $\alpha$  parameter of the H&T model
- **HT.Bet** [ $B \times (D - 1) \times \text{nBoot}$ ]  $\beta$  parameter of the H&T model
- **HT.Mu** [ $B \times (D - 1) \times \text{nBoot}$ ]  $\mu$  parameter of the H&T model
- **HT.Sig** [ $B \times (D - 1) \times \text{nBoot}$ ]  $\sigma$  parameter of the H&T model
- **HT.NonStat** [ $4 \times 1$ ] non-stationary parameter flag

- `HT.Rsd` [ $B \times 1$ ] cell array of residuals sampled for each bootstrap
- `HT.Thr` [ $B \times D - 1$ ] H&T threshold used for each of the  $D - 1$  associated variables
- `HT.NEP` [ $2 \times 1$ ] non-exceedence probability range on (0,1)
- `HT.nBoot` [scalar] number of bootstraps
- `HT.X` [ $N \times (D - 1) \times nBoot$ ] conditioning variable transformed to standard scale
- `HT.Y` [ $N \times (D - 1) \times nBoot$ ] associated variable transformed to standard scale
- `HT.RVSml` structure populated structure containing conditional return value simulations from the H&T model
- `HT.RVMth` [scalar] conditional return value method 1) random draw from the T-year max distribution 2) fixed value by bin
- `HT.RVFix` [ $B \times R$ ] for method 2, the fixed value definiton
- `HT.Sml` structure populated structure containing simulations from the H&T model
- `HT.Sml.nRls` [scalar] number of simulations from the H&T model
- `HT.Sml.Orig`: [ $(B + 1) \times RV.nRls \times D \times R$ ] simulated return values for responses on original scale
- `HT.Sml.StnMrg`: [ $(B + 1) \times RV.nRls \times D \times R$ ] simulated return values for responses on standard scale (Laplace margins)
- `HT.Sml.Unif`: [ $(B + 1) \times RV.nRls \times D \times R$ ] simulated return values for responses on uniform scale (Uniform margins)
- `Ht.Alg` [string] algorithm used to fit the H&T model either fminsearch or newton raphson, default is the latter
- `HT.n` [scalar] number of data observations  $N$
- `HT.nDmn` [scalar] number of variables (main and associated)  $D$
- `HT.SmpLclRsdOn` [boolean] flag for use of residual-sampling from local bin
- `HT.kNearest` [scalar] resampling in the k nearest neighbours for when residual-sampling from local bin is true
- `HT.ResampleDistScale` [scalar] resample distance scaling factor (written in terms of precision) from local bin is true
- `HT.nAlp` [scalar] number of  $\alpha$  parameters in the model
- `HT.nPrm` [scalar] total number of H&T model parameters
- `HT.Prm_A` [ $N \times nBoot \times 4$ ] link between H&T model parameters
- `HT.nBin` [scalar] number of covariate bins
- `HT.nRtr` [scalar] number of return periods
- `HT.A` [ $N \times nBoot$ ] matrix containing the bin allocation of the data samples in each bootstrap
- `HT.RsdInd` [ $nBoot \times 1$ ] cell containing indices of the bootstrap samples
- `HT.CVMth` [boolean] cross-validation method used (see associated entry in *Inputs* for more detail)
- `HT.nCV` [scalar] number of cross-validation groups
- `HT.nSmth` [scalar] number of smoothness parameters used

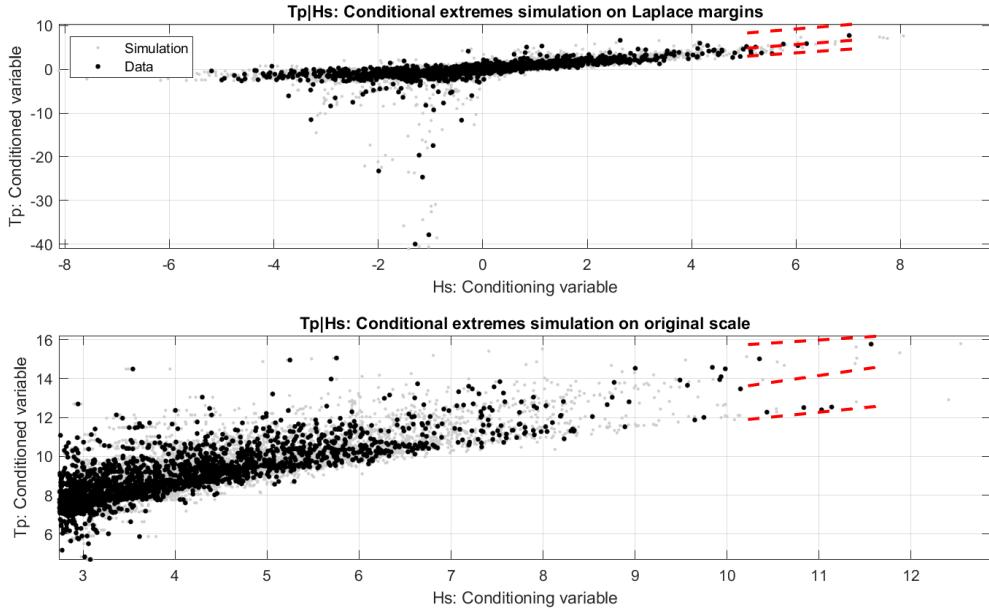


Figure 16: Comparison of original data (black points) and simulation from fitted H&T model for  $10 \times$  period of the data (grey points) on standard margins (top plot) and original scale (bottom plot). Extrapolation from the respective H&T model is shown by the red lines (5, 50, 95th percentile). On the original scale two different spikes can be seen at the upper right corner reflecting different marginal characteristics in Tp.

- HT.SmthLB [scalar] lower bound for set of candidate smoothness penalties
- HT.SmthUB [scalar] upper bound for set of candidate smoothness penalties
- HT.SmthSet [ $1 \times nSmth$ ] set of candidate smoothness penalties
- HT.OptSmth [ $1 \times nBoot$ ] optimal smoothness penalty resulting from cross-validation
- HT.CVlackOffFit [ $nSmth \times nBoot$ ] lack of fit for roughness estimation
- HT.Delta [scalar] shape parameter of the generalised Gaussian distribution (default is 2)
- HT.MarginType [string] margin type for transformation to standard scale (default is Laplace)
- HT.MinBet [scalar] minimum value for the  $\beta$  parameter of the H&T model (default is -0.5)
- HT.PrmInd [boolean] index of the number of parameters for example  $\alpha=1, \beta=2, \mu=3, \sigma=4$  changes depending on which parameters are non-stationary
- HT.AlphaLimit [ $2 \times 1$ ] limits for the  $\alpha$  parameter of the H&T model (default is  $[-1.2, 1.2]$ )
- HT.TwoParamFlag [boolean] flag for whether to fit a two parameter ( $\alpha$  and  $\beta$ ) H&T model

Figure 16 shows a comparison of the data (black points) and a simulation from the H&T model (grey points). If the simulated grey points do not reflect the distribution of the original data in black, consider reworking the data inputs to improve the model fit.

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 Tp.

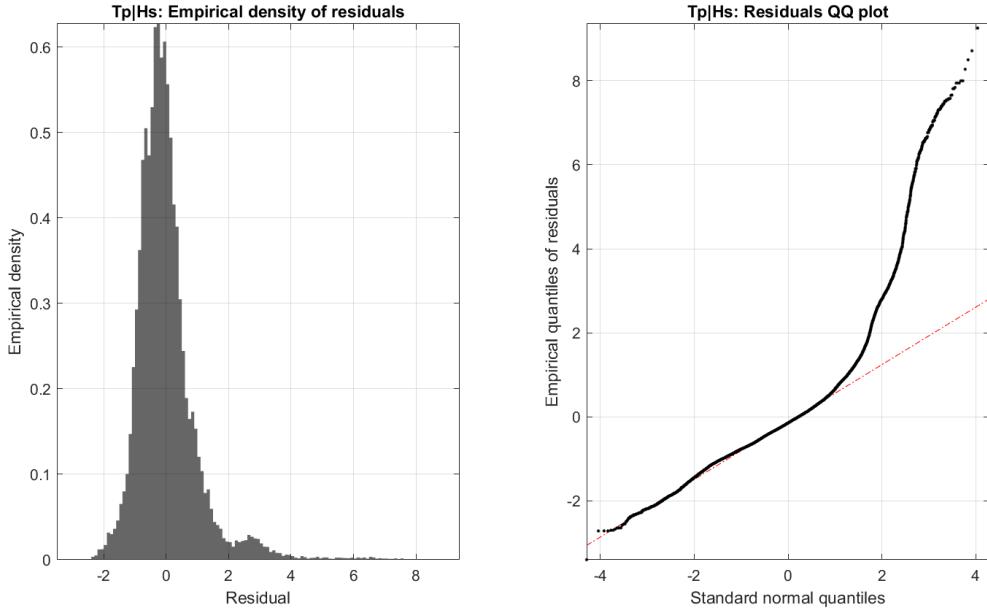


Figure 17: Diagnostic of the residuals from the H&T model fitting. Left panel shows a histogram of the residuals and the right panel shows a normal QQ plot. The right panel shows residuals as a function of direction. It is typical that these residuals are quite skewed (not normal), which is why they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

Figure 17 summarises the residuals, but this time not breaking out by covariate. The residual distribution in the left subplot is compared against the normal distribution in the right subplot by way of a normal QQ plot.

Plots of model parameter estimates and residual distributions as a function of threshold and covariate (direction) aid assessment of model-quality. Any inhomogeneity with respect to direction in Figure 18 suggests that the H&T model has not fitted well. In this case, you should adjust the H&T NEP range and possibly return to Stage 2 to adjust the covariate bin-edges to improve your representation of non-stationarity with respect to the covariate(s). The tails of the residual distribution right panel will typically not be well-matched to the line  $y = x$  (i.e. not be normally distributed) as assumed when fitting the conditional model. This is why we resample from the residuals in the simulation procedure.

The parameter themselves are presented in Figure 19 with only the  $\alpha$  parameter being a function of covariates, in this case direction. For the remaining parameters histograms are shown to reflect the uncertainty across the different bootstraps.

The threshold stability plot in Figures 20 is 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 H&T NEP. A range of [0.7,0.85] would seem to be a reasonable choice here.

Figure 21 shows the return value CDFs for the North Sea Tp|Hs example. Here the omni-directional CDF is bimodal, this is largely due to directional differences in the Tp marginal distribution. Similar effects can be seen in Figure 16.

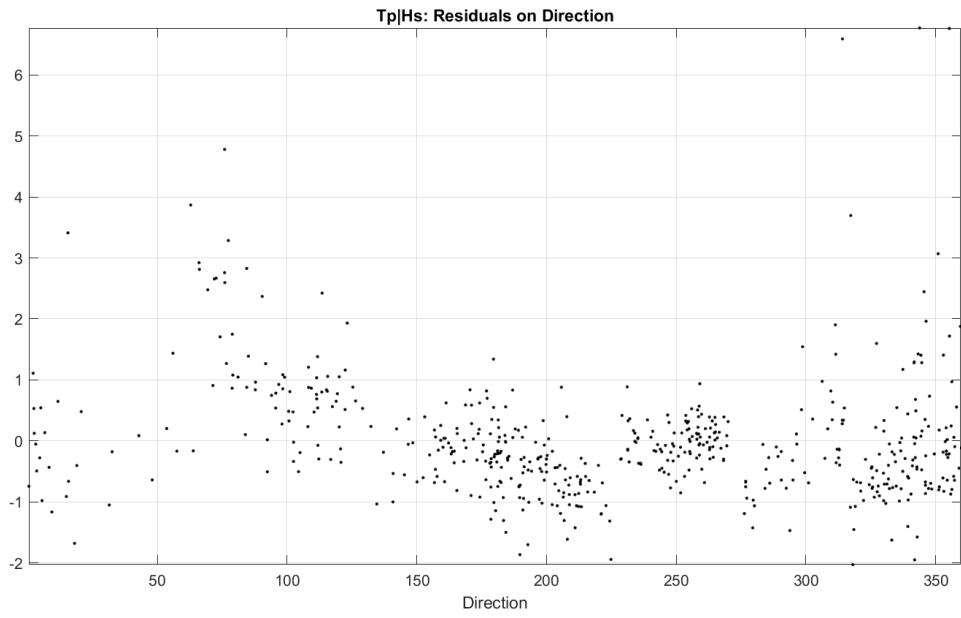


Figure 18: Diagnostic of the residuals from the H&T fitting in the presence of covariates in this case direction which are plotted against the residuals of the H&T model.

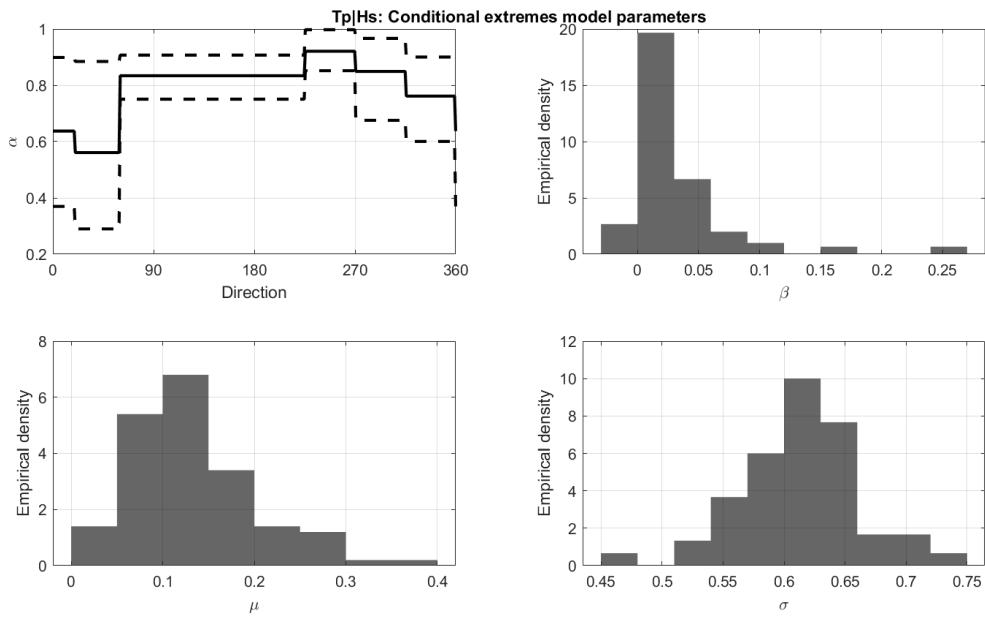


Figure 19: Histograms of the H&T parameters over bootstrap re-samples in a non-stationary case. The parameter uncertainty captures, marginal (bootstrap and NEP) and conditional (bootstrap and H&T NEP) uncertainty.

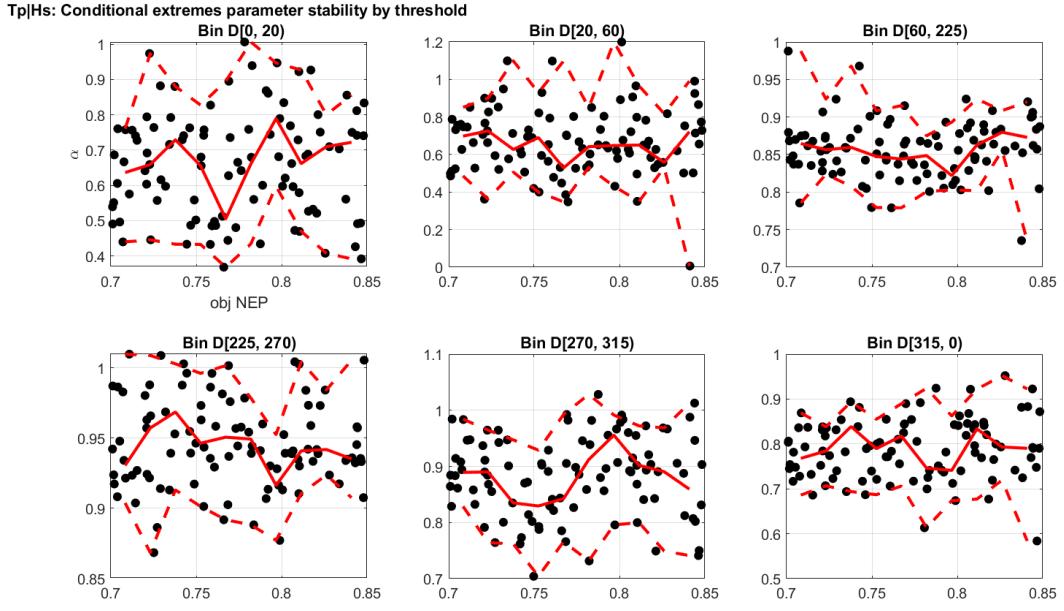


Figure 20: H&T parameter  $\alpha$  as a function of the H&T NEP. Black dots show individual bootstrap estimates, red lines are local binned median, 2.5 and 97.5 percentile estimates. A well behaved model should be stable over a range of NEP's.

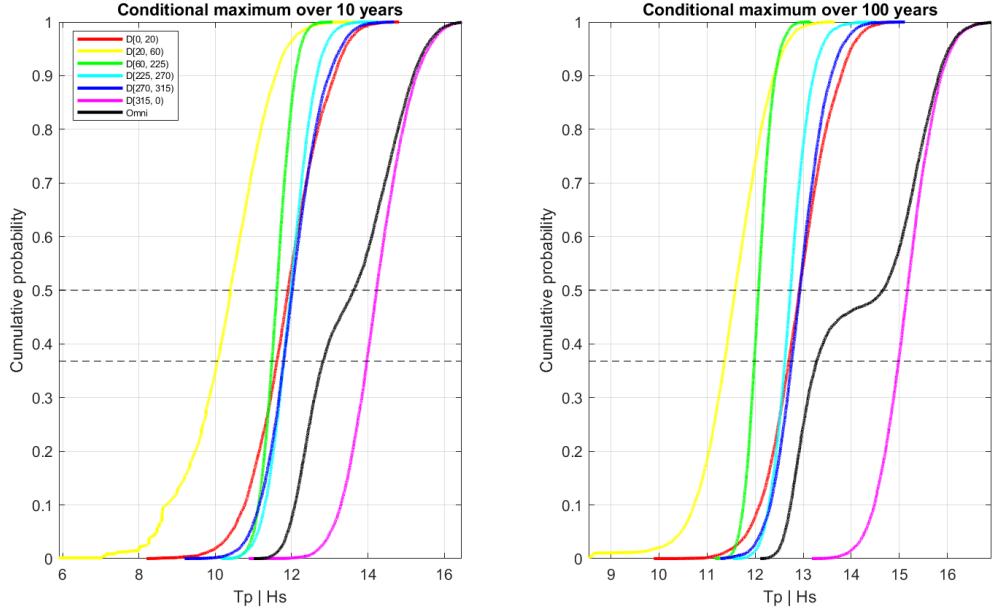


Figure 21: Conditional return value CDF's  $p(Tp|Hs_{10})$  and  $p(Tp|Hs_{100})$ . Directional CDFs are shown using coloured lines. Black line shows the omni-directional estimate. The two reference levels at 0.368 and 0.5 that correspond to the most probable quantile ( $\exp(-1)$ ) and the median quantile. In this case using the North Sea data the omni-directional CDF is bimodal, this is largely due to marginal differences in Tp.

## 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] 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  $Y_d$  (for any  $d \in \{2, \dots, D\}$ ). To efficiently compute the contours, a new importance sampling method was written into the covXtreme software.

The **Constant Exceedance Contour** preserves the observation count in the extremal set, as illustrated in Figure 24. The region of simultaneously extreme  $Y_1$  and  $Y_d$  is captured using 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, the other contour methods do not preserve probabilities in this way. From a risk point of view it can therefore be deemed to be robust.

The **Huseby Contour** is 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. Fast changing parts of the contour can come out ‘spiky’ so we also apply a moving average to try to overcome some of these issues.

- Downside: The convexity doesn’t behave well in multimodal cases or other cases where the data is non-convex;
- Upside: A complete contour is produced - covering the full angular space around its centre point, not just the region with the largest responses.

The **Heffernan and Tawn density** contour is defined by gridding the data on the original scale and calculating the density of simulations in each bin, and then drawing the line which preserves the density of the bin containing the lock point. Figure 25 illustrates this method. The ‘omni’ contour is computed as a weighted sum across the covariate-binned contours (weighted by rate of occurrence).

- Downside: the contour is non-invariant to transformations of variables;
- Upside: we get a contour which hugs the data in the way we might expect (without a table top, and without the extra roundness sometimes induced by the convexity assumption of the Huseby contour).

### 5.1 Running Stage5

Marginal and conditional model parameters are loaded from Stages 3 and 4 respectively. The inputs are specified as part of OptionsContour as `OptCnt`. We recommend only changing the recommended inputs to change below when initially estimating the contours. If there are issues with the estimation of the contours we then

**Run Scripts:** `Stage5_Condition.m`

**Output files:** `Output\Cnt.mat`

**Recommend inputs to change:**

- `OptCnt.Mth` [string] cell array of contour methods to be used:
  - `Exc`: constant exceedance contour
  - `Hus`: convex “Direct Sampling” contour of Huseby
  - `HTDns`: constant density contour on standard margins, uses density form of H&T to get contour

- `OptCnt.nSml` [scalar] 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

**Additional inputs:**

- `OptCnt.nGrd` [scalar] number of grid points for each of  $Y_1$  and  $Y_D$  when rectangular gridding needed
- `OptCnt.nPnt` [scalar] number of points on the contour
- `OptCnt.SmtWdtC` [scalar] smoothing width for the
- `OptCnt.BndWdtScl` [scalar] bandwidth scale for the Heffernan and Tawn density contour

**Output:** Contour structure `Cnt`

- `Cnt.nPnt` [scalar] how many points from which to draw contour
- `Cnt.Rng` [`nPon`  $\times$   $(K + 1) \times R$ ] conditioned values for contour, for  $K$  covariate bins and  $R$  contour levels
- `Cnt.XY` [ $nMth \times 1$ ] cell array for the contour lines. In case of `Exc` and `Hus` methods  $XY(i)$  is [`nPon`  $\times 2 \times (B + 1) \times R \times (D - 1)$ ] defining contour lines in case of `HTDns`  $XY(i)$  is a  $[(B + 1) \times R \times (D - 1)]$  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.PltOn` boolean flag to switch on exploratory diagrams
- `Cnt.LvlOrg`  $[(B + 1) \times R \times (D - 1)]$ , contour level on original scale of conditioned variable (lock point)

Figures 22 and 23 show comparisons of the 3 methods for the North Sea  $T_P|H_S$  example.

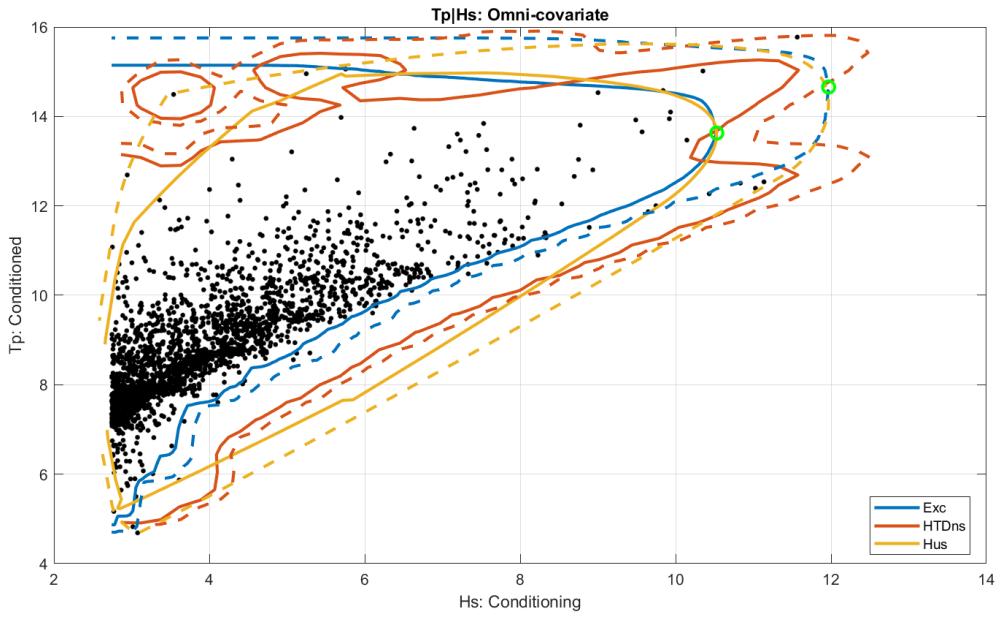


Figure 22: Comparison of contour methods omni-directionally. Contours are for 10 and 100 year return periods (solid and dashed lines respectively). Different methods are shown using coloured lines. The green circle shows the lock point at each probability level through which all the contour methods have to pass.

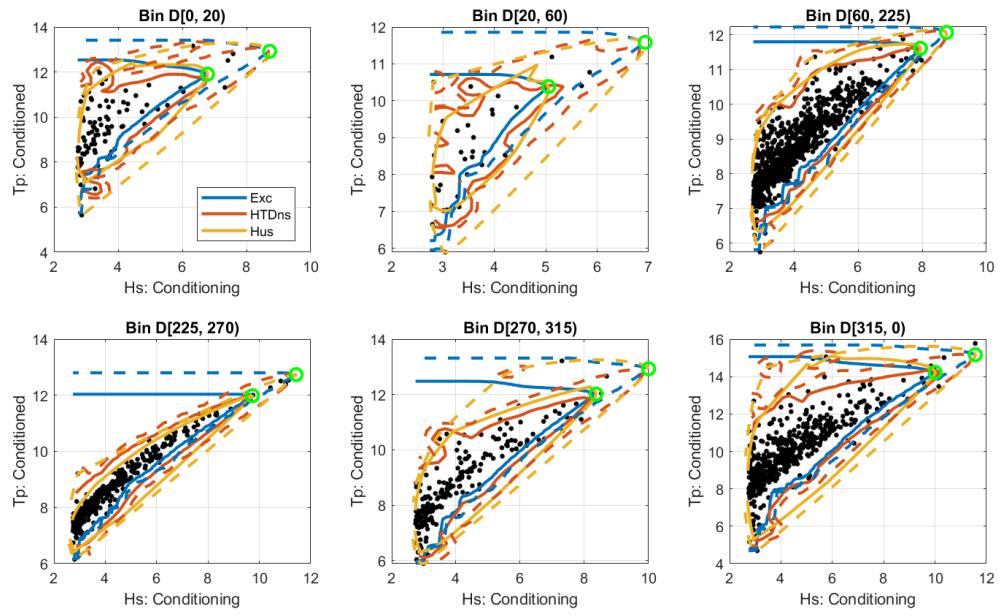


Figure 23: Comparison of contour methods by directional sector. Contours are for 10 and 100 year return periods. Different methods are shown using coloured lines. The green circle shows the lock point at each probability level through which all the contour methods have to pass.

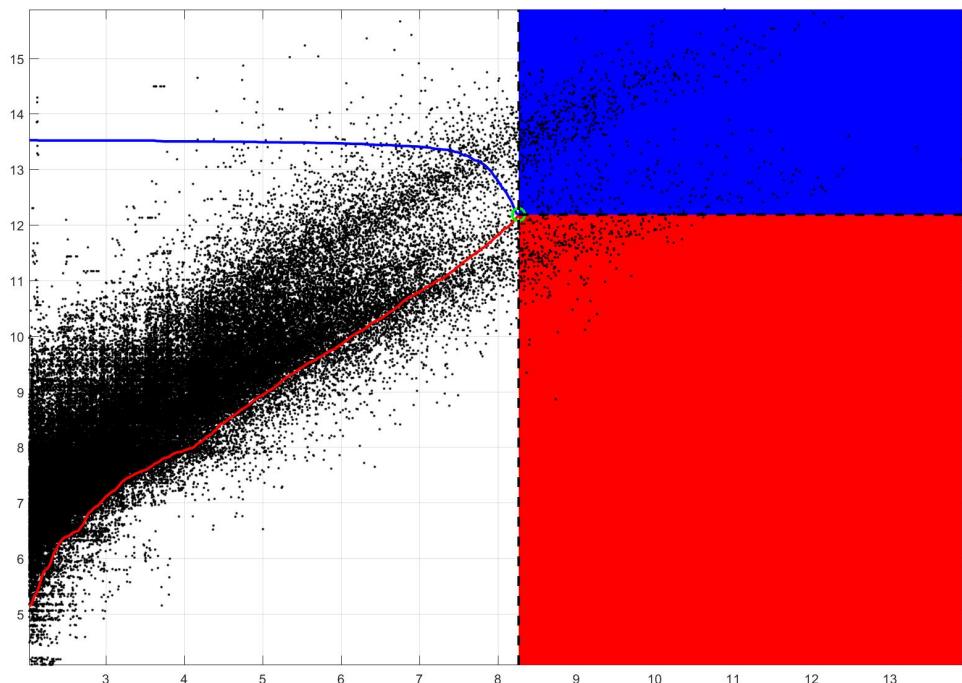


Figure 24: Illustration of constant exceedance contour. Green lock point is defined using the extreme quantile in  $Y_1$  and the conditional median in  $Y_2$ . The blue line is drawn such that, as  $Y_1$  is decreased, the number of observations in the blue quadrant is preserved. The red line is drawn such that, as  $Y_1$  decreases, the number of observations in the red quadrant is preserved.

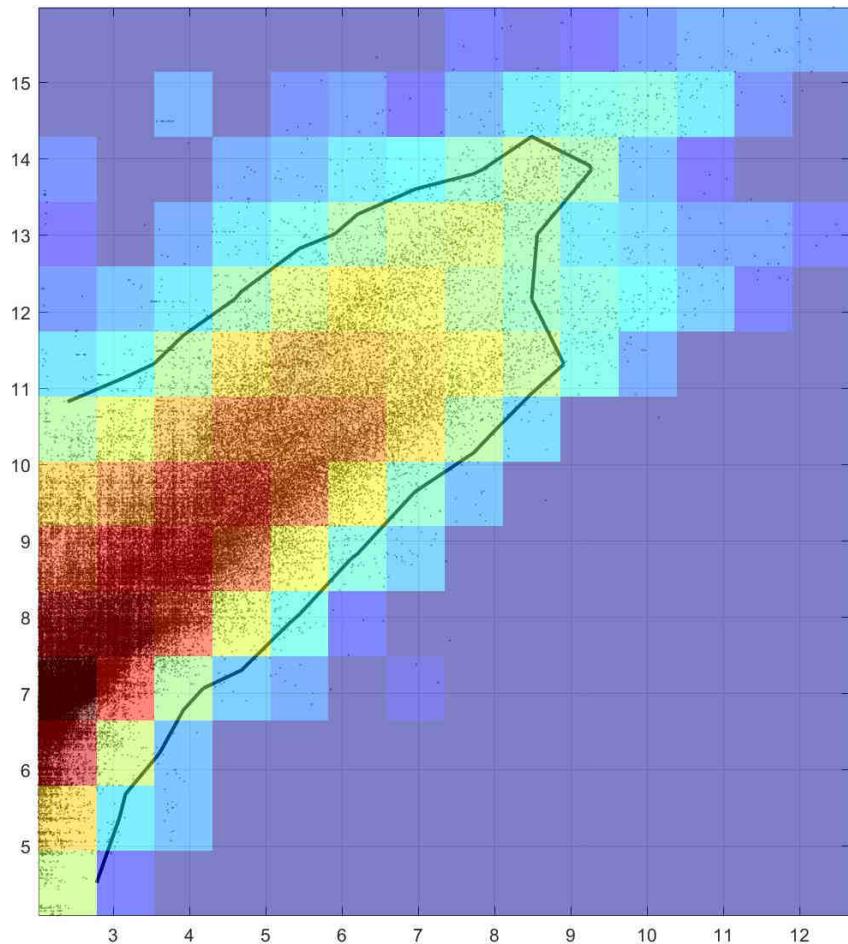


Figure 25: Illustration of empirical density contour. The coloured squares represent the observation counts per bin.

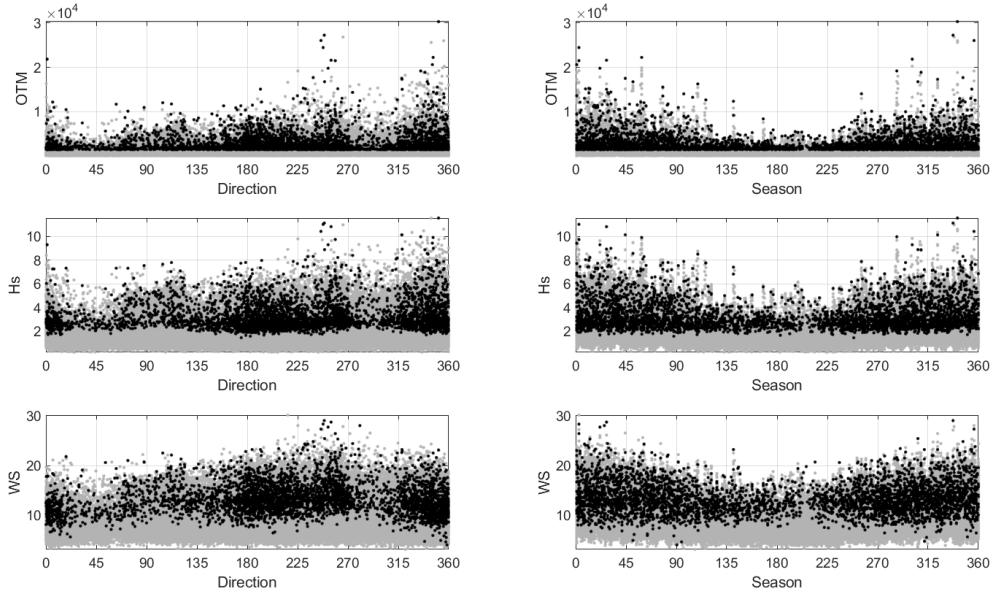


Figure 26: Scatter plots of variables (black points) and their storm peak values (grey points) by covariate - in this case direction and season. Each row of subplots relates to a different variable (OTM, Hs and WS respectively), and each column to a different covariate.

## 6 Example: Multiple Associated Variables and Covariates

Having illustrated the key inputs and outputs of the code for a simple case - a single associated variable and a single covariate - in this section we provide a higher-dimensional example (you'll find the associated MATLAB run-scripts in the `CaseStudy` folder). We set the main (conditioning) response to over-turning moment (OTM) and use two associated variables - significant wave height (Hs) and wind speed (WS). Further, we will use two periodic covariates: wave direction and season. The resulting figures below illustrate the result of running the covXtreme software with higher dimensions.

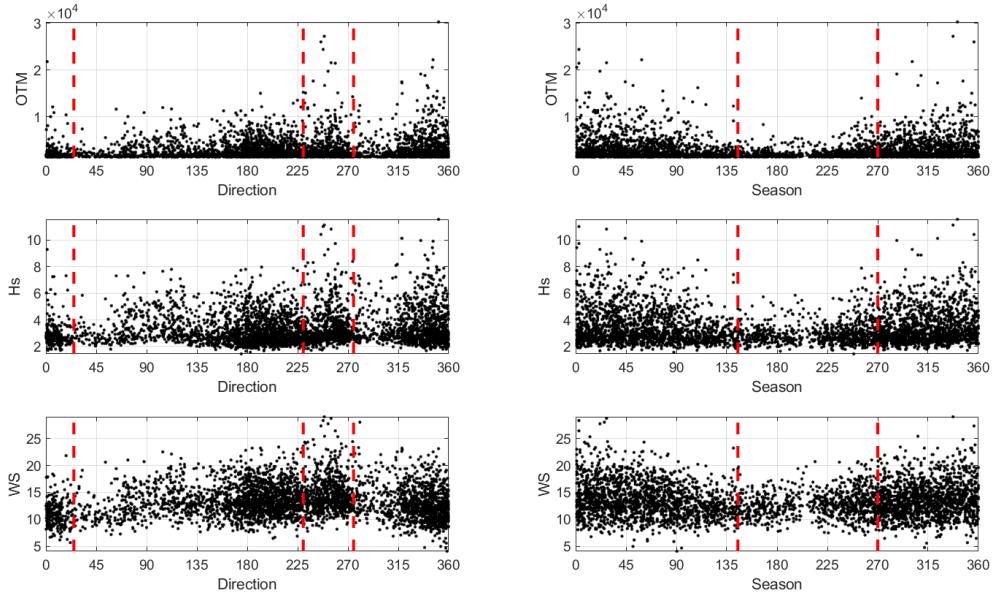


Figure 27: Example bin allocation. Bins are chosen at  $[25, 230, 275] \times [145, 270]$  degrees in direction and season. Storm peak data shown in black, chosen bin edges are shown with red dashed lines. Each row of subplots relates to a different response; and each column to a different covariate.

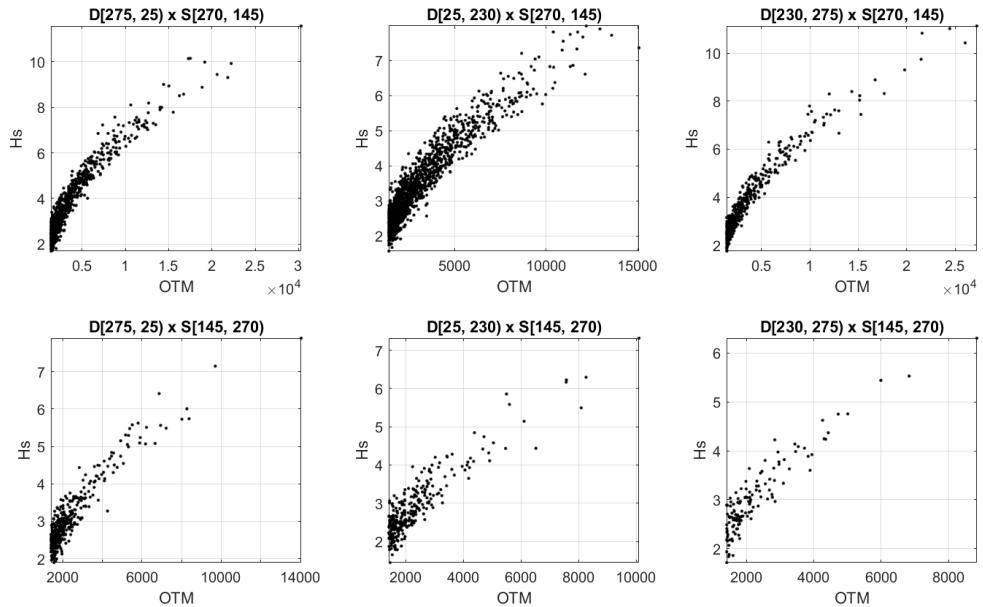


Figure 28: Pairwise scatter plots of storm peak data for an associated variate ( $Y_2=Hs$ ) on the conditioning variate ( $Y_1=OTM$ ), partitioned by covariate bin (different panels). Equivalent figure files exist for each associated variate. In this example there are  $3 \times 2$  combinations of directional-seasonal covariate bins.

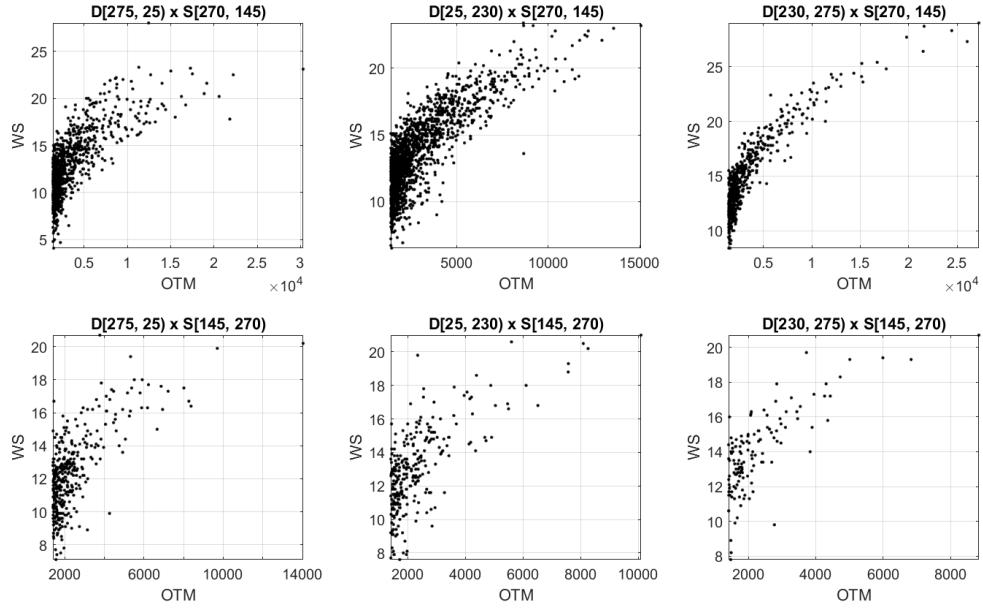


Figure 29: Pairwise scatter plots of storm peak data for an associated variate ( $Y_3=WS$ ) on the conditioning variate ( $Y_1=OTM$ ), partitioned by covariate bin (different panels). Equivalent figure files exist for each associated variate. In this example there are  $3 \times 2$  combinations of directional-seasonal covariate bins.

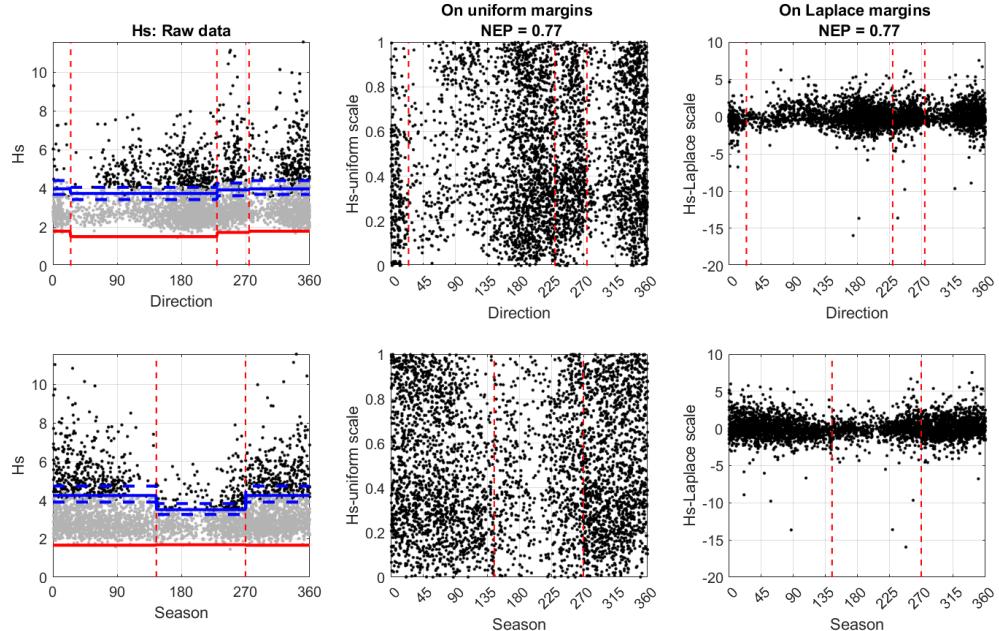


Figure 30: Left column shows sea-state data for a *single response* in grey and storm peaks in black plotted against direction (top row) and season (bottom row). Bin edges are indicated by dashed red lines. 2.5, 50 and 97.5 percentiles of estimated threshold across all bootstraps are plotted with blue lines. The gamma location parameter is plotted in solid red. The middle and right columns show the response data on uniform and standard scale. An equivalent figure file exists for each conditioned response.

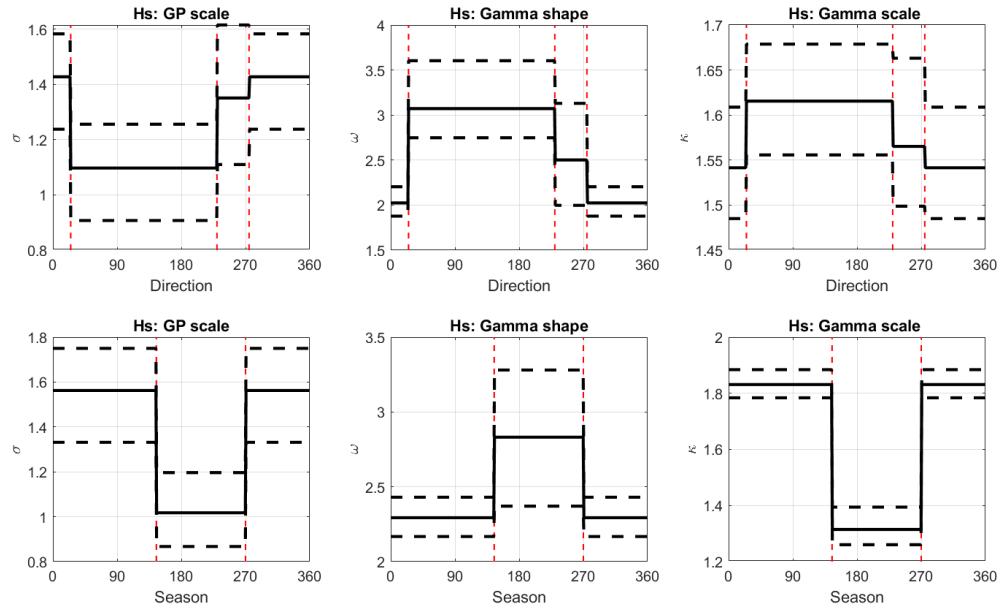


Figure 31: Black lines show 2.5, 50 and 97.5 percentiles of GP and Gamma parameters as a function of direction (first row) and season (second row). Red lines show bin edges. Each column refers to a different parameter: GP scale; Gamma shape and scale. An equivalent figure file exists for each variate  $Y_d$ ,  $d = 1, \dots, D$ .

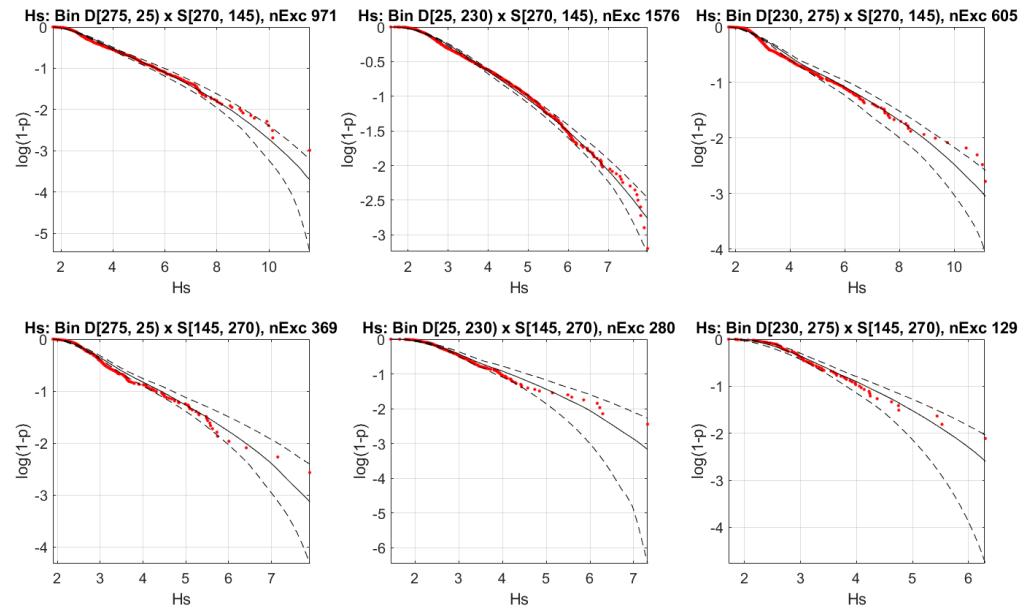


Figure 32: Empirical and model-based tails. Diagnostic for quality of model-fit by covariate bin. Red dots show storm peaks, black lines are 2.5, 50 and 97.5 percentiles of bootstrap uncertainty band under the estimated model. Red dots inside the confidence limits of the model indicate good fit.

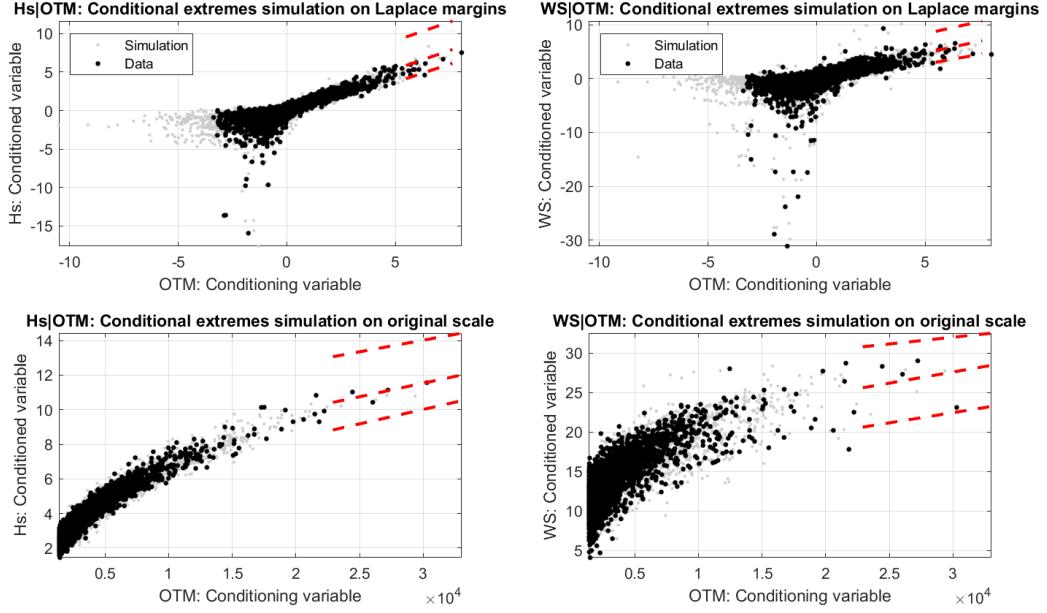


Figure 33: Comparison of original data (black points) and simulation from fitted H&T model for  $10 \times$  period of the data (grey points) on standard margins (top row) and original margins (bottom row). The red points show the extrapolation lines from the respective H&T models with the median and the 5 and 9th quantile. Each column relates to a different associated variate ( $Y_d, d > 1$ ; Hs and WS in this case) conditioned on the conditioning variate  $Y_1$  (OTM), with  $D = 3$  in this case.

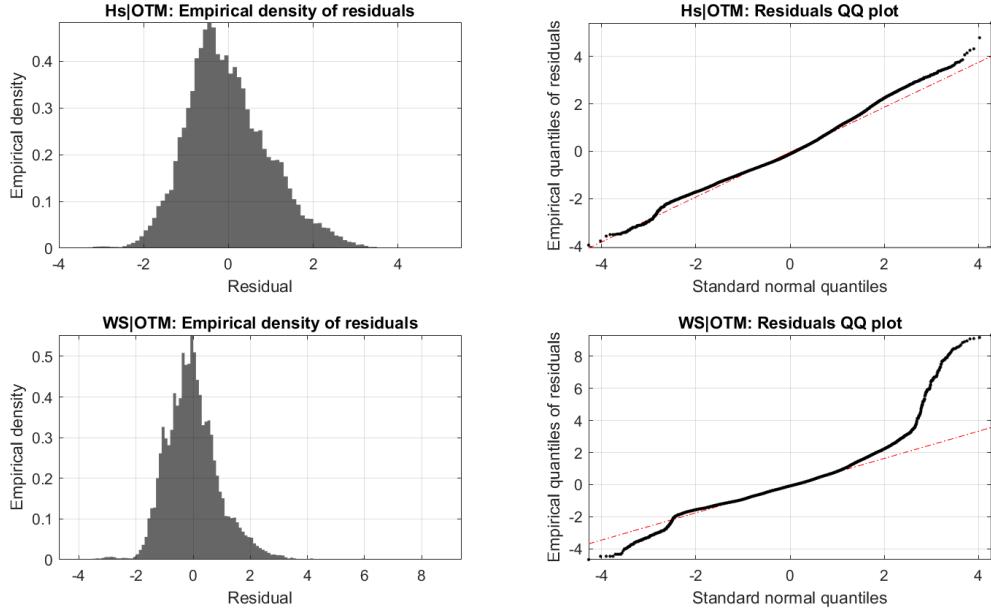


Figure 34: Diagnostic of the residuals from H&T model fitting, where each row of plots corresponds to a different associated variable ( $Y_d, d > 1$ ) conditioned on  $Y_1$ . Column 1 contains histograms of the residuals. The second column compares this distribution with the normal distribution (the form assumed in the model) via a normal QQ plot. It is typical that the tails of the distribution of residuals are quite long (i.e. not normal), which is why they are reused in the simulation procedure.

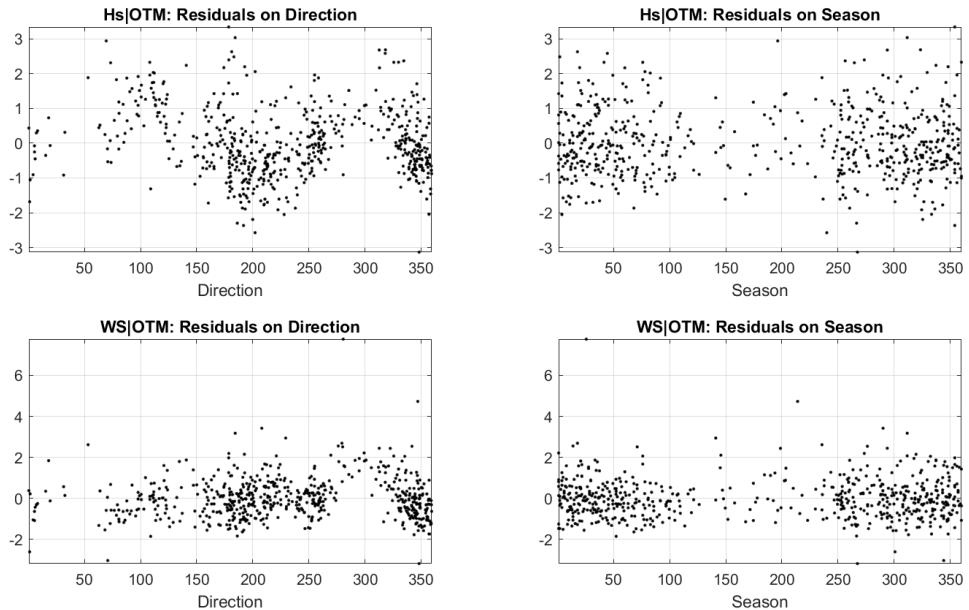


Figure 35: Diagnostic of the residuals from H&T model fitting: residuals plotted against covariate(s). Row  $d - 1$  corresponds to the model  $Y_d|Y_1$  for  $d \in \{2, \dots, D\}$ . Each column shows the residuals plotted against a different covariate: first column direction and second column season. Inhomogeneity of residuals with respect to a covariate (assessed using this plot) points to a model which does not sufficiently account for covariate effects, in which case the user should reconsider their covariate bin choice and/or NEP choice(s). Note an outlier for WS|OTM at about 270 degrees, which occurs early on in the year.

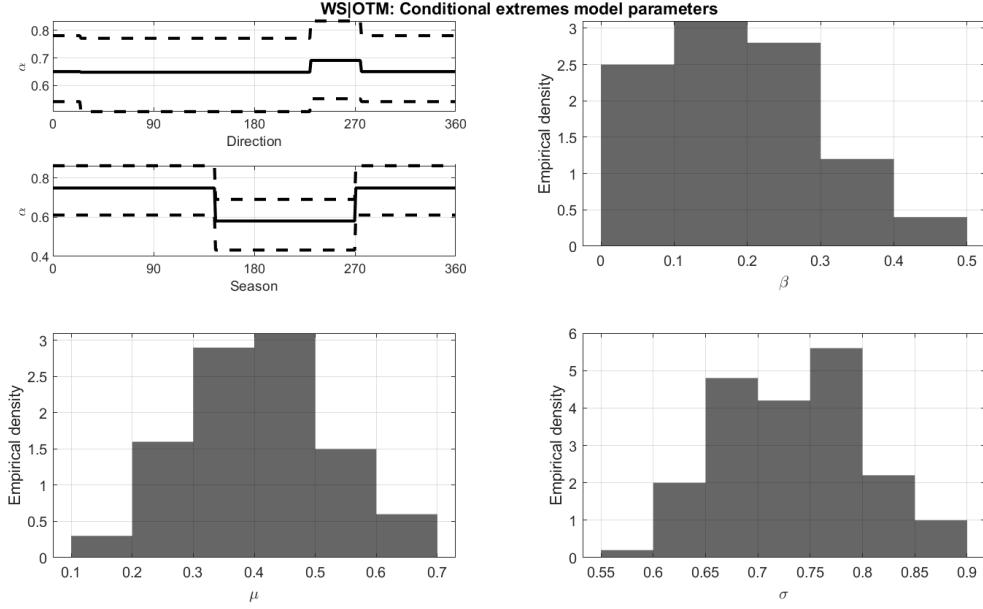


Figure 36: Plots characterising the bootstrap uncertainty of H&T parameters for WS|OTM. In this case a non-stationary H&T model was fitted, meaning that the top left panel contains plots of the  $\alpha$  parameter as a function of the 2 covariates, direction and season. The solid black line shows the parameter estimate changing as a function of bin with the dashed line showing the associated uncertainty in the parameter estimate. The remaining panels contain histograms of the H&T parameters  $\beta$  (top right),  $\mu$  (bottom left) and  $\sigma$  (bottom right) respectively, which are specified not to vary by covariate bin (either direction or season). The parameter uncertainty captures marginal (bootstrap and NEP) and conditional (bootstrap and H&T NEP) uncertainty. An equivalent figure is available for each of the associated H&T models.

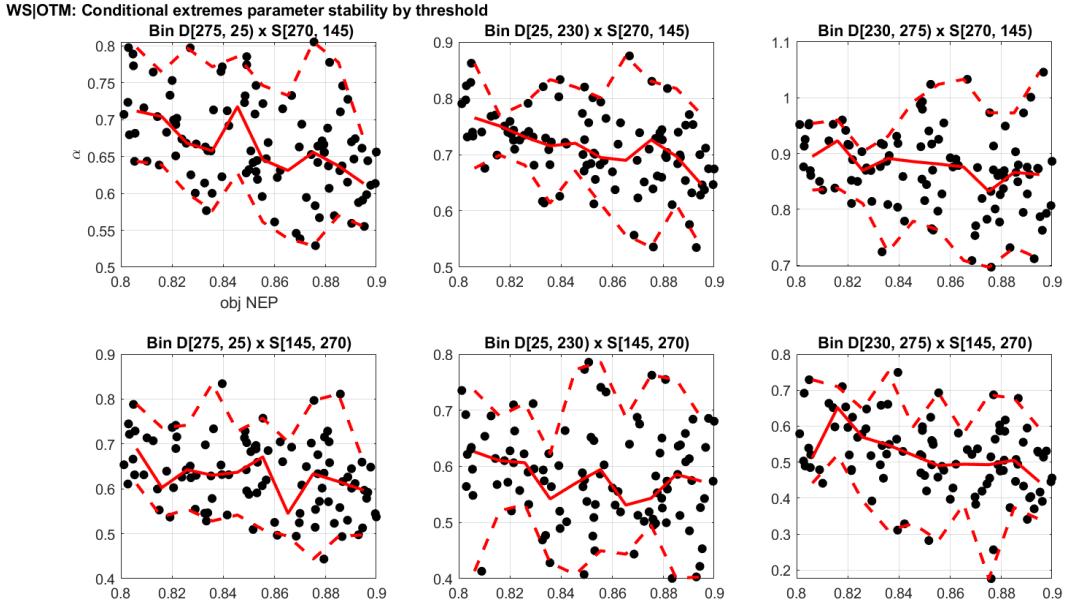


Figure 37: H&T parameter  $\alpha$  as a function of the H&T NEP for the model WS|OTM, each panel corresponds to a different directional-seasonal combination. Black dots show individual bootstrap estimates, red lines are local binned median, 2.5 and 97.5 percentile estimates. A well behaved model should be stable over a range of NEP's. An equivalent figure is available for each of the  $Y_d|Y_1$  models.

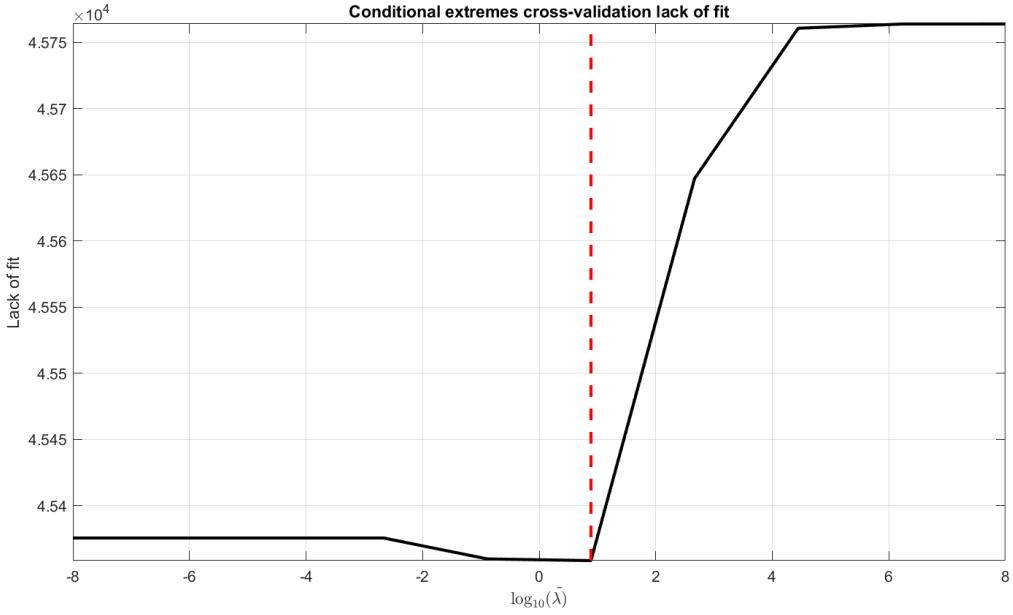


Figure 38: Estimation of lack of predictive fit against chosen smoothness  $\lambda$  of the H&T parameters. Low indicates good predictive performance. The red line indicates the optimal choice which is the minimum value of the lack fit, the scale is misleading as the values of  $\lambda$  between -8 and 1 are similar to one another. When the red line is at the edge of the  $\lambda$  domain, the user should extend the range of  $\lambda$  considered using the inputs HT.SmthLB and HT.SmthUB.

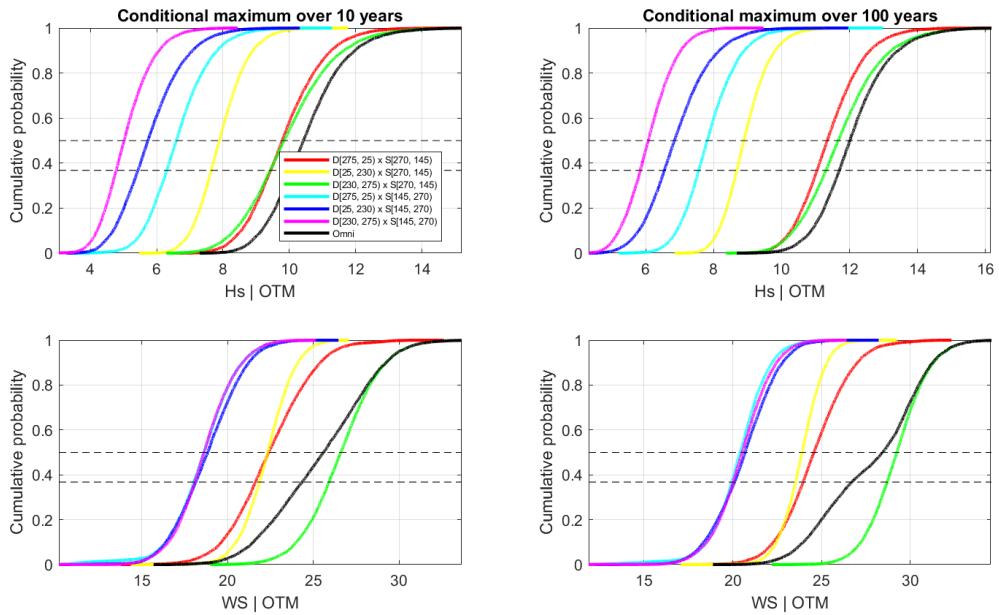


Figure 39: Conditional 10 (first column) and 100 (second column) return value CDF's of  $p(Hs|OTM)$  and  $p(WS|OTM)$ . CDFs for each covariate bin are shown using coloured lines. Black line shows the omni-directional estimate.

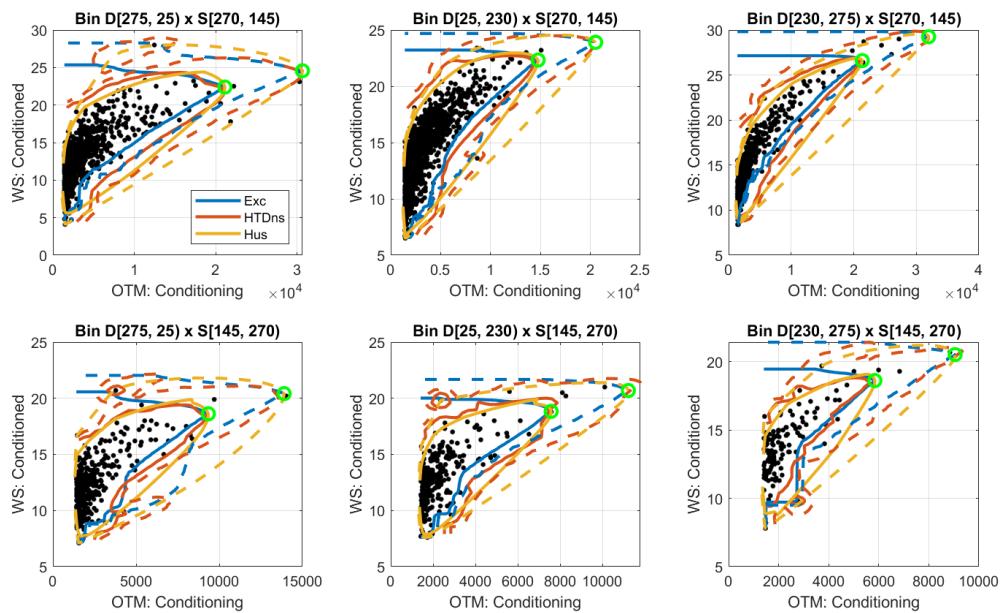


Figure 40: Contour plots for a single associated response (conditioned on the main response), in this case  $WS|OTM$ , broken out by covariate bin. Two contour lines are drawn: one for the 10 year and one for the 100 year return period. Each coloured line corresponds to a different contouring method. The green circle shows the lock point at each probability level through which all the contour methods have to pass. The dimension of the subplot array flexes with the number of covariates and covariate bins. An equivalent figure file exists for each of the  $D - 1$  associated responses.

## A Appendix: Model details

This section contains details of the underlying statistical models used in the covXtreme software. For marginal modelling a generalised Pareto distribution is used above a pre-determined threshold and below a truncated gamma distribution. For the case of dependence modelling we adopt the conditional extremes model of Heffernan and Tawn (2004). The implementation of these models in the code are discussed in more detail in Sections 3 and 4 respectively.

Let  $\dot{\mathbf{Y}} = (\dot{Y}_1, \dots, \dot{Y}_D)$  be a set of  $D$ -dimensional random variable corresponding to a set of responses on their original margin (otherwise known as their physical scale). The values of  $\dot{Y}_1$  are used to define the storm peaks over the data with associated values of  $\dot{Y}_2, \dots, \dot{Y}_D$ . For the set of responses, we have a set of  $n$  independent observations  $(\dot{\mathbf{y}}_1, \dots, \dot{\mathbf{y}}_n)$  along with associated covariates  $\{\mathbf{x}_i\}_{i=1}^N$  on some domain  $\mathcal{X}$ . We apply the model as discussed in Section A.1 to each margin to transform the data to common scale, in this case Laplace margins represented as  $\mathbf{Y} = (Y_1, \dots, Y_D)$ . The procedure for this marginal transformation is explained in Section A.2. Through obtaining the set of  $\mathbf{Y}$  we are able to model the joint behaviour of the values conditional on  $Y_1$  being large. We adopt the conditional extremes model which is presented in Section A.3, this allows us to capture different extents of dependence of the variables in  $\mathbf{Y}$ .

Non-stationarity with respect to covariate bins is captured via the GP scale and the parameters of the conditional extremes model all of which can vary across the bins. The extent of variation across these bins is controlled via a roughness-penalty in the marginal and conditional extreme value likelihood functions with more details found below.

### A.1 Marginal model

For each variable  $\dot{Y}_d$ ,  $d = 1, 2, \dots, D$ , and covariate bin  $b$ ,  $b = 1, 2, \dots, B$ , we independently estimate a three-parameter gamma distribution using maximum likelihood estimation with sample likelihood

$$\ell_{\text{Gmm}}(\omega_{bd}, \kappa_{bd}, l_{bd}; \{\dot{y}_{id}\}_{i=1}^N) = \prod_{b=1}^B \prod_{i: A(i)=b} f_{\text{Gmm}}(\dot{y}_{id}; \omega_{bd}, \kappa_{bd}, l_{bd})$$

for sample  $\{\dot{y}_{id}\}_{i=1}^N$ , where  $f_{\text{Gmm}}$  is the density of the gamma distribution with shape  $\omega_{bd} \in \mathbb{R}$ , scale  $\kappa_{bd} > 0$  and location  $l_{bd} \in \mathbb{R}$  given by

$$f_{\text{Gmm}}(\dot{y}_{id}; \omega_{bd}, \kappa_{bd}, l_{bd}) = \frac{\kappa_{bd}^{-\omega_{bd}}}{\Gamma(\omega_{bd})} (\dot{y}_{id} - l_{bd})^{\omega_{bd}-1} \exp\left(\frac{\dot{y}_{id} - l_{bd}}{\omega_{bd}}\right).$$

where  $\Gamma(\cdot)$  is the gamma function. We adopt the 3-parameter gamma distribution, which is an orthogonal parameterisation taken from Cox and Reid [1987]. For visualisation of the parameters of the gamma distribution, we revert back to the scale and shape representation of the gamma distribution.

Subsequently, we calculate the extreme value threshold  $\psi_{bd}(\tau_d) = F_{\text{Gmm}}(\tau_d; \hat{\omega}_{bd}, \hat{\kappa}_{bd}, \hat{l}_{bd})$  using the estimated gamma parameters, for the pre-specified non-exceedance probability  $\tau_d$ , where  $F_{\text{Gmm}}$  is the cumulative distribution function of the gamma distribution. The marginal sample likelihood for threshold exceedances of  $\psi_{bd}$  for variable  $\dot{Y}_d$  over all covariate bins can therefore be written

$$\ell_{\text{GP}}(\xi_d, \{\nu_{bd}\}_{b=1}^B; \{\dot{y}_{id}\}_{i=1}^N, \{\psi_{bd}\}_{b=1}^B) = \prod_{b=1}^B \prod_{i: A(i)=b; \dot{y}_{id} > \psi_{bd}(\tau_d)} f_{\text{GP}}(\dot{y}_{id}; \xi_d, \nu_{bd}, \psi_{bd})$$

with the first product over covariate bins and the second product over the set of observations within the given covariate bin. The  $f_{\text{GP}}$  is the density of the generalised Pareto distribution with shape  $\xi_d \in \mathbb{R}$  and scale  $\nu_{bd} > 0$  given by

$$f_{\text{GP}}(\dot{y}_{id}; \xi_d, \nu_{bd}, \psi_{bd}) = \frac{1}{\nu_{bd}} \left[ 1 + \frac{\xi_d}{\nu_{bd}} (\dot{y}_{id} - \psi_{bd}) \right]^{-1/\xi_d - 1}.$$

Recall from Section 3 that we have an additional parameter, the smoothness penalty  $\lambda_d$ , which controls the extent to which the GP shape  $\nu_{bd}$  can vary across bins. Its optimal value,  $\hat{\lambda}_d$ , is chosen to maximise predictive performance (defined as optimally regulating the extent to which  $\nu_{bd}$  can vary across the

bin) using a k-fold cross-validation procedure. This results in the following roughness-penalised negative log-likelihood function  $\mathcal{L}_{GP}$ :

$$\begin{aligned}\mathcal{L}_{GP}(\xi_d, \{\nu_{bd}\}_{b=1}^B, \lambda_d; \{\dot{y}_{id}\}_{i=1}^N, \{\psi_{bd}\}_{b=1}^B) &= -\ell_{GP}(\xi_d, \{\nu_{bd}\}_{b=1}^B, \lambda_d; \{\dot{y}_{id}\}_{i=1}^N, \{\psi_{bd}\}_{b=1}^B) \\ &\quad + \lambda_d \left( \frac{1}{B} \sum_{b=1}^B \nu_{bd}^2 - \left[ \frac{1}{B} \sum_{b=1}^B \nu_{bd} \right]^2 \right).\end{aligned}\quad (1)$$

Parameters are estimated to minimise the penalised negative log likelihood for each response in turn, with the GP and gamma parameters carried forward to subsequent inference being those associated with the optimal  $\hat{\lambda}_d$  resulting from cross-validation. Note that there is no penalisation for the gamma likelihood below the threshold  $\psi_{bd}$ .

We have likelihood functions of this form for each variate  $d$ , meaning that each variate will come with a different set of fitted parameters. Also that the spatial location of bins in multiple covariates (i.e. their relative location on a grid) does not factor into the roughness-penalty. For example, a case with 4 directional bins and 3 seasonal bins still boils down to a likelihood which is a product over the total  $B = 4 \times 3 = 12$  bins.

## A.2 Marginal transformation to standard scale

The marginal models are used to transform from the original scale to a standard scale (Laplace) using the probability integral transform. For legacy reasons, we also present the transformation to standard Gumbel margins. The following marginal transformation is presented for an example bin  $b$  (for simplicity we drop the bin subscript), and each bin is transformed in turn along with its associated parameters to capture any non-stationarity.

1. First we transform data above the marginal threshold  $\tau$  to uniform scale using the GP cumulative distribution function (CDF):

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

for  $\dot{y} \in (\psi, \dot{y}^+]$  where  $\dot{y}^+$  is the upper end point of the distribution. This is also rescaled to account for the probability that an observation  $\dot{y}$  is above the threshold  $\psi$ .

Below the threshold, a gamma CDF is used:

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

where  $\gamma(\cdot, \cdot)$  is the lower incomplete gamma function. Through doing this marginal transformation we obtain a set of uniformly distributed data  $u$ .

2. Given the uniformly distributed  $u$ , we can then transform to standard Laplace margins by using the inverse of the standard Laplace CDF:

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

where  $\text{sgn}$  is the sign function. For reference, this is derived from the standard Laplace CDF:

$$F_L(z) = \mathbf{1}(z) - 0.5 \text{sgn}(z) \exp(-|z|),$$

where  $z \in \mathbb{R}$  and  $\mathbf{1}(z)$  is an indicator function defined as follows

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

An equivalent process is also to transform the uniform data  $u$  to standard **Gumbel** margins by using the inverse of the standard Gumbel CDF:

$$y = F_G^{-1}(u) = -\log(-\log(u)).$$

This is derived from the Gumbel CDF:

$$F_G(z) = \exp(-\exp(-z)),$$

where  $z \in \mathbb{R}$ .

### A.3 Conditional extremes model of Heffernan and Tawn (2004)

The marginally transformed Laplace-scale  $d$ -dimensional sample  $\mathbf{Y} = (Y_1, Y_2, \dots, Y_D)$  is modelled by using the conditional extremes model of Heffernan and Tawn (2004). We consider values of the variate  $Y_1$  above a dependence threshold  $\phi(\tilde{\tau}) = F_L^{-1}(\tilde{\tau})$ , for which the conditional extremes model holds and the parameters of the model can be estimated. The model is defined as followed for all  $y_1 > \phi(\tilde{\tau})$

$$(Y_2, Y_3 \dots Y_D) | (Y_1 = y_1) = \boldsymbol{\alpha}_{\tilde{\tau}b} y_1 + y_1^{\beta_{\tilde{\tau}b}} \mathbf{W}_{\tilde{\tau}b},$$

with parameters  $\boldsymbol{\alpha}_{\tilde{\tau}b} = (\alpha_{\tilde{\tau}b2}, \dots, \alpha_{\tilde{\tau}bD}) \in [-1, 1]^{D-1}$  and  $\boldsymbol{\beta}_{\tilde{\tau}b} = (\beta_{\tilde{\tau}b2}, \dots, \beta_{\tilde{\tau}bD}) \in [-\infty, 1]^{D-1}$ .

The working assumption is made for model estimation that the residual distribution  $\mathbf{W}_{\tilde{\tau}b} \sim \text{GG}(\boldsymbol{\mu}_{\tilde{\tau}b}, \text{diag}(\boldsymbol{\sigma}_{\tilde{\tau}b}^2), \boldsymbol{\delta})$  with mean  $\boldsymbol{\mu}_{\tilde{\tau}b} \in \mathbb{R}^{D-1}$  and variance  $\boldsymbol{\sigma}_{\tilde{\tau}b} \in [0, \infty)^{D-1}$ . The GG stands for the generalised Gaussian distribution (otherwise known as the delta-Laplace distribution) and the  $\boldsymbol{\delta} = (\delta_2, \dots, \delta_d)$  parameter controls the shape of the distribution with  $\boldsymbol{\delta} = 1, 2$ . For model fitting, the distribution of the residuals  $\mathbf{W}_{\tilde{\tau}b}$  are initially assumed to be independent of the conditioning variable  $Y_1$  and for estimation we assume that the covariance matrix of  $\mathbf{W}_{\tilde{\tau}b}$  is diagonal; this allows us to fit each of  $D - 1$  pairs of random variables in turn, for example  $(Y_2 | Y_1 = y_1), \dots, (Y_D | Y_1 = y_1)$  for  $y_1 > \phi(\tilde{\tau})$ . After model estimation, in order to maintain the dependence between  $Y_2, \dots, Y_D$  we jointly resample from the empirical residual distribution of  $\mathbf{W}_{\tilde{\tau}b}$ . For simplicity of notation we drop the indexing of threshold  $\tilde{\tau}$ , however it should be noted that the parameters of the conditional extremes model are of course dependent on the choice of threshold.

If we consider a pair of random variables  $Y_1$  and  $Y_d$  for which the associated conditional extremes model has been fitted for a particular bin  $b$ , the model has the following set of residuals  $W_{bd}$  with marginal density

$$f_{W_{bd}}(w; \mu_{bd}, \sigma_{bd}, \delta_d) = \frac{\delta_d}{2\kappa(\delta_d)\sigma_{bd}\Gamma(\frac{1}{\delta_d})} \exp\left\{-\left|\frac{w - \mu_{bd}}{\kappa(\delta_d)\sigma_{bd}}\right|^{\delta_d}\right\},$$

where  $\kappa(\delta_d)^2 = \Gamma(1/\delta_d)/\Gamma(3/\delta_d)$ . If  $\delta_d = 1(2)$  the residual distribution corresponds to a Laplace(Gaussian) distribution. Within the code the user has the ability to choose the value of  $\delta_d$  for each of the different conditioned responses.

The mean ( $m_{bd}$ ) and standard deviation ( $s_{bd}$ ) of the conditional distribution  $Y_d | (Y_1 = y_1)$  are as follows  $m_{bd} = \alpha_{bd}y_1 + \mu_{bd}y_1^{\beta_{bd}}$  and  $s_{bd} = y_1^{\beta_{bd}}\sigma_{bd}$ . The parameters of the conditional extremes model are estimated to minimise the following negative log-likelihood  $\tilde{\ell}$ , with the  $\sim$  symbol used to denote a dependence likelihood rather than a marginal likelihood as given in Section A.1:

$$\tilde{\ell}(\boldsymbol{\alpha}_d, \boldsymbol{\beta}_d, \boldsymbol{\mu}_d, \boldsymbol{\sigma}_d; \mathbf{y}_d, \tilde{\tau}, \delta_d) \propto \sum_{b=1}^B \sum_{\substack{i; b=A(i) \\ y_{di} > \phi(\tilde{\tau})}} \left[ \log(\sigma_{bd}) + \beta_{bd} \log(y_{i1}) + \left| \frac{y_{id} - \alpha_{bd}y_{i1} - \mu_{bd}y_{i1}^{\beta_{bd}}}{\kappa(\delta_d)\sigma_{bd}y_{i1}^{\beta_{bd}}} \right|^{\delta_d} \right]. \quad (2)$$

for all  $y_1 > \phi(\tilde{\tau})$ . In the case where we want the  $\boldsymbol{\alpha}_d$  parameter of the conditional extremes model to change within each bin, we add an extra term to the likelihood given in (2), this is the same setup as for the marginal model in (1) but centered around the parameter. Further terms would be added to the penalised negative log likelihood  $\tilde{\mathcal{L}}$  in (3) if we introduced non-stationarity into the other parameters of

the conditional extremes model. However, we recommend focussing on incorporating non-stationarity into the  $\alpha_d$  parameter with the associated as follows

$$\tilde{\mathcal{L}}(\boldsymbol{\alpha}_d, \boldsymbol{\beta}_d, \boldsymbol{\mu}_d, \boldsymbol{\sigma}_d, \tilde{\lambda}_d; \mathbf{y}_d, \tilde{\tau}, \delta_d) = -\tilde{\ell}(\boldsymbol{\alpha}_d, \boldsymbol{\beta}_d, \boldsymbol{\mu}_d, \boldsymbol{\sigma}_d; \mathbf{y}_d, \tilde{\tau}, \delta_d) + \tilde{\lambda}_d \left( \frac{1}{B} \sum_{b=1}^B \left[ \alpha_{bd} - \frac{1}{B} \sum_{b=1}^B \alpha_{bd} \right]^2 \right) \quad (3)$$

for each value of  $\tilde{\tau}$ .

Once estimates of the parameters of the conditional extremes model have obtained, the associated residuals  $r_{id}$  (we drop the bin index  $b$  for readability) can be calculated. The residuals for a given response  $d$  are defined as follows:

$$r_{id} = \frac{1}{\hat{\sigma}_{dy_{1d}}^{\hat{\beta}_d}} \left( y_{id} - \hat{\alpha}_d y_{i1} - \hat{\mu}_d y_{i1}^{\hat{\beta}_d} \right) \text{ for all } i \text{ such that } b = A(i) \text{ and } y_{i1} > \phi(\tilde{\tau}).$$

The residuals are then jointly resampled (thereby preserving the dependence between them) to make predictions rather than simulating from the assumed residual distribution used for model estimation, which assumed conditional independence between the residuals. A similar process also exists below the dependence threshold,  $\phi(\tilde{\tau})$ , whereby the data are jointly resampled.

## 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.