covXtreme: open-source software for modelling extreme environment data sets

Ross Towe, Emma Ross, David Randell and Philip Jonathan

Motivation

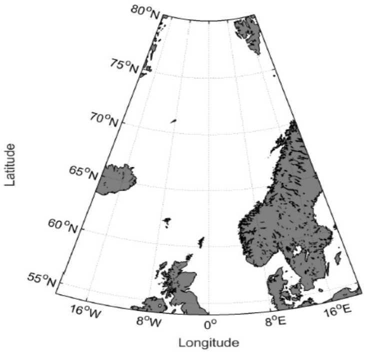
- Offshore operations **require the probability of failure** of manned structures and ships to be at the level of p=1e-4 per annum. The so called in **1 in 10 000 year event**.

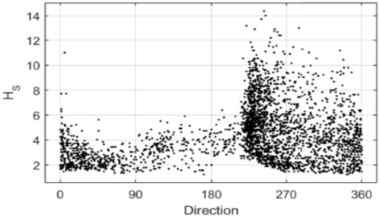
- This requires the understanding of the natural environment:
  - Extreme weather
  - Joint behaviour of waves, winds and currents
  - Impact of covariates such as direction and the time of year

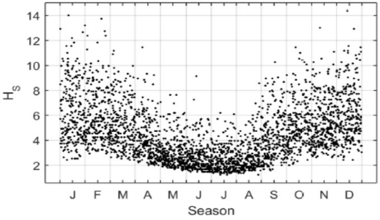
- Want to be able to **propagate and quantify uncertainty** related to modelling oceanographic data

Motivation

Oceanographic data







Motivation

- Statistical tool should handle the following features:
  - Accurate estimation of the tails of a data set
  - Capture covariate effects such as direction and season
  - Account for the interaction between multiple variables
  - Careful handling of uncertainty

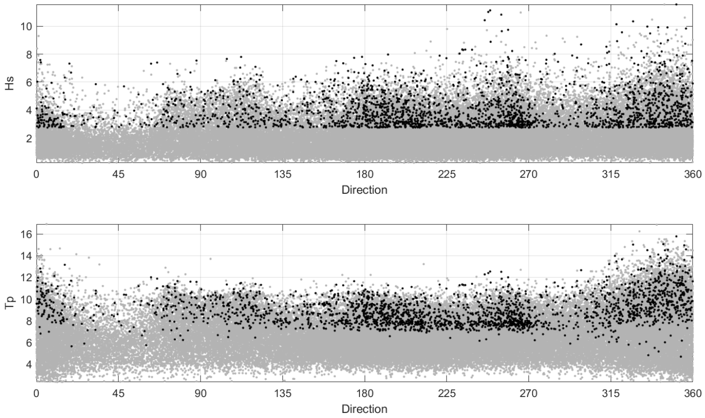
- As a result, we have developed covXtreme, a open source MATLAB software for the estimation of extreme environmental conditions.

- Previous example applications of the code include Ross et al. [2017], Ross et al. [2020], Guerrero et al. [2021] and Barlow et al. [2023]

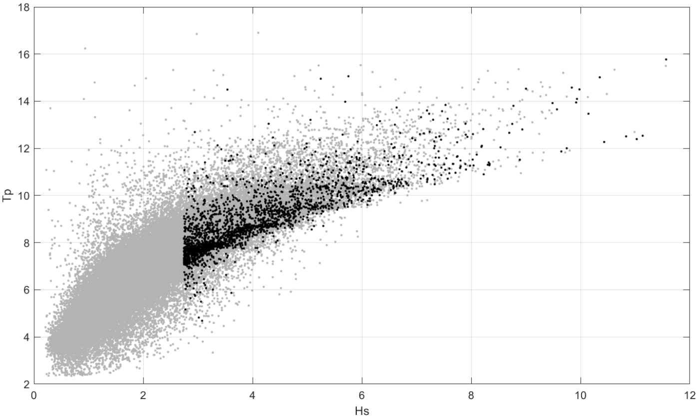
covXtreme

- **Stage 1:** selection of extreme events from an enivronmental data sets or simulation of a data set: selection of independent events
- **Stage 2:** selection of covariate bins, for example wave height as a function of direction: capture covariates for upcoming marginal modelling
- **Stage 3:** estimation of marginal models with respect to covariates: non-stationary modelling as a function of bin
- **Stage 4:** joint estimation of oceanographic variables, for example the behaviour of wind speed when wave height is large: account for interaction between multiple varibales
- **Stage 5:** estimation of environmental contours for risk assessment: intrepretable summary for design engineers
- Example modelling the relationship between significant wave height (Hs) and peak period (Tp) - included in the user guide

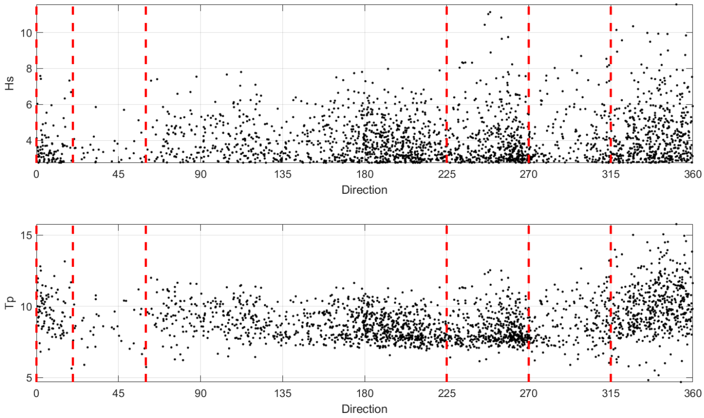
Stage 1: extraction of storm peaks



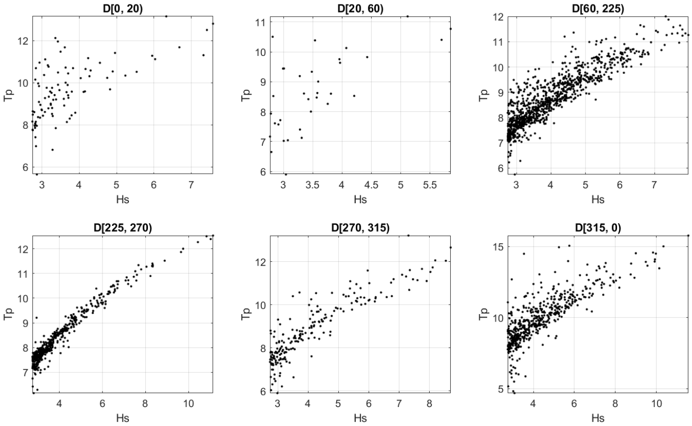
Stage 1: extraction of storm peaks



Stage 2: selection of bins



Stage 2: joint behaviour of Hs and Tp



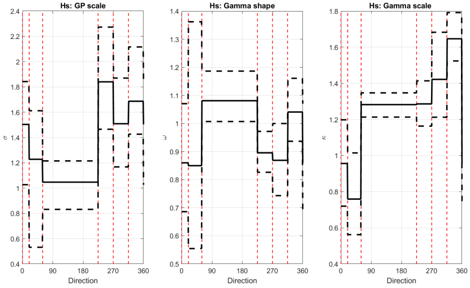
Stage 3: marginal model

- Set a bin dependent threshold ψ_b to define extreme events
- For data below the threshold fit a **Gamma distribution**
- For data above the threshold fit a **generalised Pareto (GP) distribution**:
    - Threshold ψ_b with scale ν_b and shape parameter ξ
- Likelihood above the threshold:

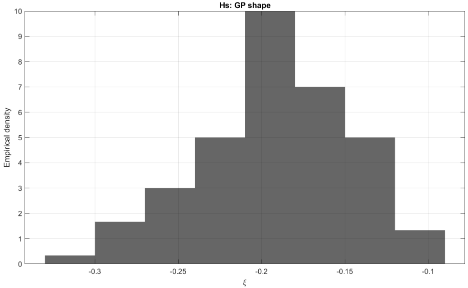
$$\ell(\dot{x}_i \mid \xi, \nu_b, \psi_b, \lambda) = \log \prod_{b=1}^{B} \prod_{\substack{i;A(i)=b;\\ \dot{x}_i>\psi_b}} f_{GP}(\dot{x}_i \mid \xi, \nu_b, \psi_b) + \lambda \left(\frac{1}{B} \sum_{b=1}^{B} \nu_b^2 - \left[\frac{1}{B} \sum_{b=1}^{B} \nu_b \right]^2 \right)$$

Stage 3: marginal model (Hs)

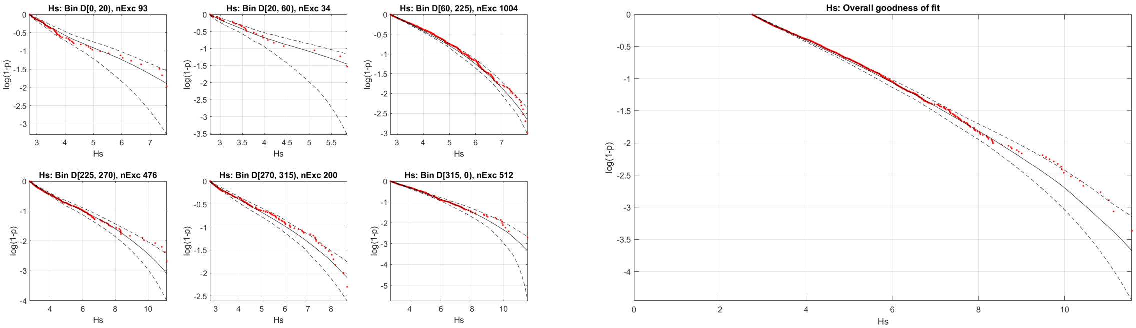
GP scale and Gamma parameters



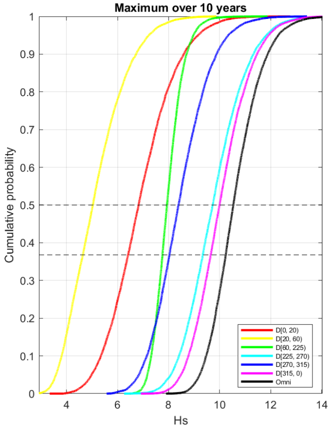
GP shape parameter

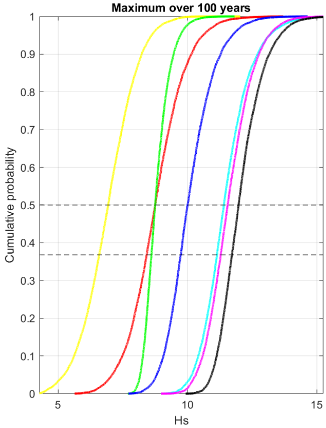


Stage 3: marginal model assessment (Hs)



Stage 3: marginal return values (Hs)



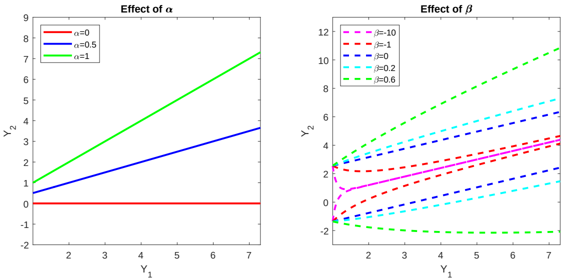


Stage 4: dependence model

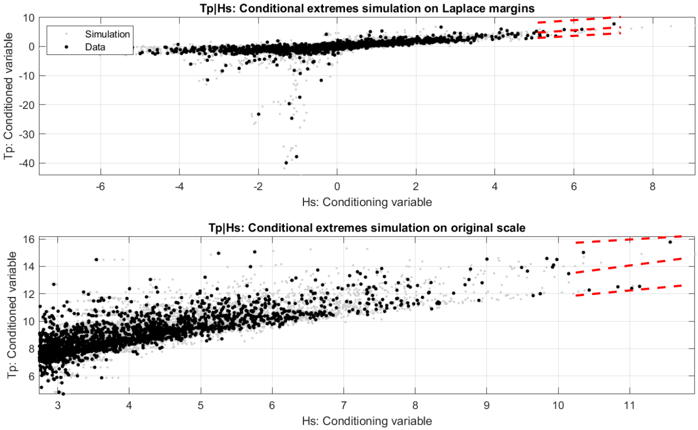
Conditional dependence model of Heffernan and Tawn [2004]:

$$(Y_2 | Y_1 = y) = \alpha_b y + y^{\beta_b} W$$

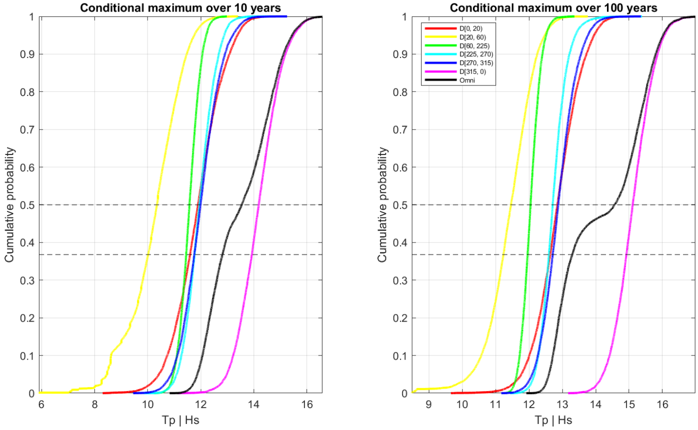
- $Y_2 = \text{Tp}$, $Y_1 = \text{Hs}$ *on Laplace scale*
- for $y >$ sufficiently large threshold ϕ
- $\alpha_b \in [-1, 1]$, $\beta_b \in (-\infty, 1]$
- $W \sim \text{DeltaLaplace}(\mu_b, \sigma_b, \delta)$



Stage 4 - simulations from the dependence model



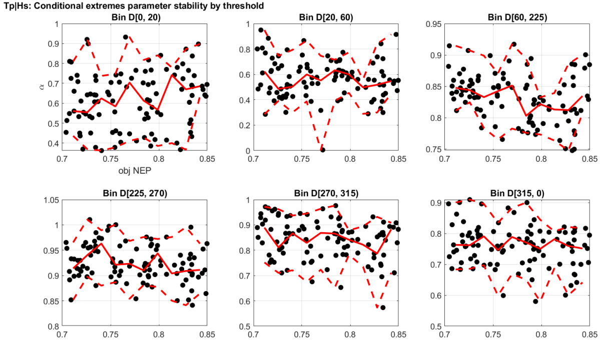
Stages 3 and 4: conditional return values (Tp|Hs)



Stages 3 and 4: dealing with uncertainty

Two sources of uncertainty:

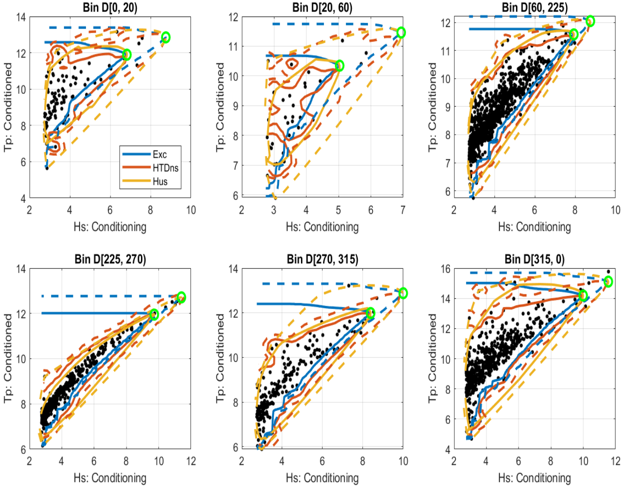
- Bootstrap resampling
- Non exceedance probability threshold: $\tau \sim \text{Unif}(\tau_{LB}, \tau_{UB})$



Dependence model threshold assessment

Stage 5: contour estimation

- Estimation of risk profiles
- Three different contour methods:
  - Exceedance (Exc)
  - Heffernan and Tawn (HTDns)
  - Huseby (Hus)
- Number of control factors



Discussion

- covXtreme code enables quick analysis of extreme environmental data sets
- Based on appropriate statistical methods for marginal and dependence modelling
- Ability to consider both observed and simulated data sets
- Ability to handle covariate information, for example direction and season
- Propagation of uncertainty
- Code will be made openly available through GitHub

References

A M Barlow, E Mackay, E Eastoe, and P Jonathan. A penalised piecewise-linear model for non-stationary extreme value analysis of peaks over threshold. *Ocean Eng.*, 2023.

Matheus B. Guerrero, Raphaël Huser, and Hernando Ombao. Conex-connect: Learning patterns in extremal brain connectivity from multi-channel eeg data, 2021. URL https://arxiv.org/abs/2101.09352.

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

E. Ross, D. Randell, K. Ewans, G. Feld, and P. Jonathan. Efficient estimation of return value distributions from non-stationary marginal extreme value models using Bayesian inference. *Ocean Eng.*, 142:315–328, 2017.

E. Ross, O. C. Astrup, E. Bitner-Gregersen, N. Bunn, G. Feld, B. Gouldby, A. Huseby, Y. Liu, D. Randell, E. Vanem, and P. Jonathan. On environmental contours for marine and coastal design. *Ocean Eng.*, 195:106194, 2020.