Estimation of extreme risks in a bivariate setting

Callum Murphy-Barltrop
STOR-i Centre for Doctoral Training, Lancaster University

Supervised by: Jennifer Wadsworth, Emma Eastoe

January 16th, 2023

Overview

Motivation

Return curves

Improved estimation of the angular dependence function

Non-stationary return curves

Discussion

Motivation

- 1999 flood at the Blayais nuclear power plant.
- Damage to the plant's off-site power supply - crucial for cooling reactors.
- Level 2 event on the International Nuclear Event Scale - "Very close to a major accident".

Motivation

- Why do we care? (Why was it interesting?)
- The flood was due to a combination of extreme winds AND sea-levels.
- Implies we cannot consider these variables in isolation when considering extreme risks.
- New methods of evaluating flood risk at nuclear sites were developed.
- Was in some sense a forerunner of the 2011 Fukushima nuclear disaster - also a combination of two extreme events (earthquake + tsunami).

Motivation

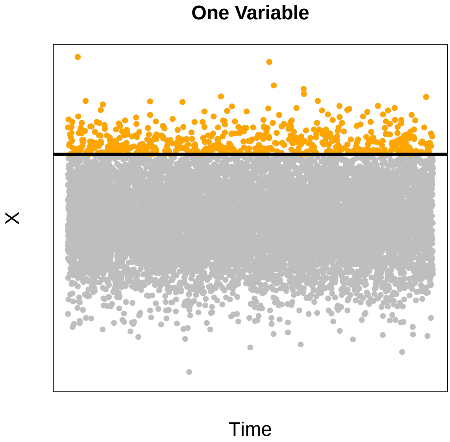
This results in the following questions:

- How do we **define** an extreme event for two (or more) variables?
- How do we **model** these events?
- How do we **summarise** the relevant **risks**?

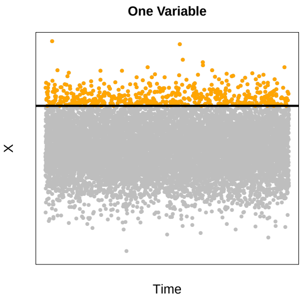
We restrict attention to the bivariate (two variable) setting.

Motivation

- What about the univariate (one variable) setting?
- This problem is well-studied.



Motivation



- Fit generalised Pareto distribution to exceedances of some high threshold (Coles, 2001).
- Use fitted model to estimate return levels: given a small probability p close to 0, return level x_p satisfies

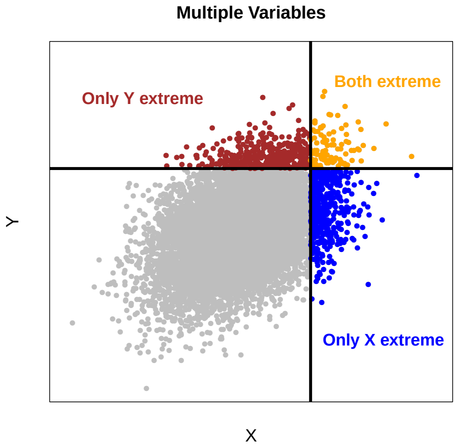
$$\Pr(X > x_p) = p.$$

Motivation

- Return levels (high quantiles) widely used to summarise extremal risks in the univariate setting.
- E.g. design of flood defences in the UK (D'Arcy et al., 2022).
- A wide range of approaches exist for their estimation.

Motivation

- What is extreme for two variables?



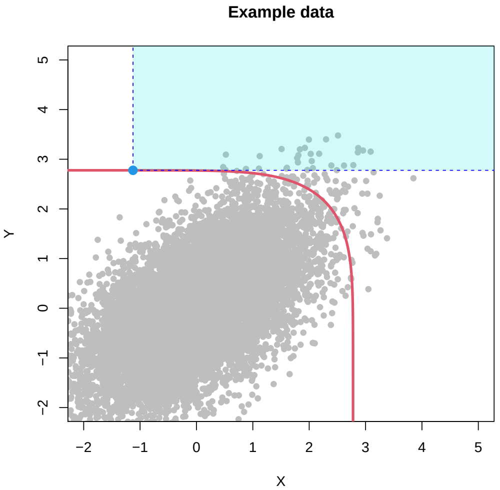
Motivation

- Need to consider extreme behaviour in all three regions.
- To do this, we expand the notion of a return level.
- Given a (continuous) random vector (X, Y), we define the set

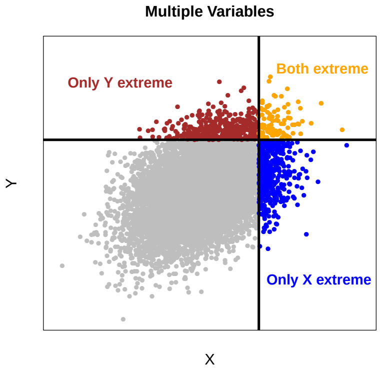
$$RC(p) := \{(x, y) \in \mathbb{R}^2 \mid \Pr(X > x, Y > y) = p\},$$

  where p is very small.
- Provides a summary of joint extremes.

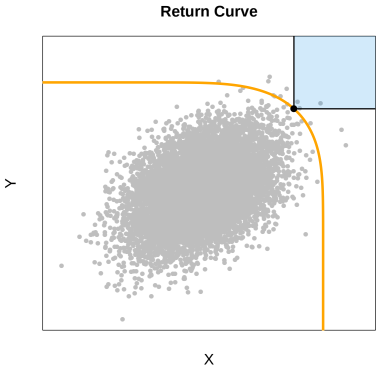
Motivation



Motivation



Multiple Variables
Only Y extreme
Both extreme
Only X extreme
Y
X



Return Curve
Y
X

Motivation

- We have our risk measure for bivariate extremes.
- But how do we estimate it?
- For very small probabilities, standard statistical techniques fail - extrapolating outside the data range.
- Motivate the use of bivariate extreme value theory.

Return curves

- Return curve estimation not been well studied.
- Motivated the first paper of my PhD, in which I considered:

1. Return curve estimation.
2. Quantifying uncertainty.
3. Evaluating goodness of fit.

Return curves - estimation

- A note on dependence/copula modelling: we assume standardised marginal distributions for random vectors (X, Y).
- In practice, this can be achieved through the probability integral transform.
- If $X \sim F_X$, $Y \sim F_Y$, then

$$(U, V) := (F_X(X), F_Y(Y)),$$

has standard uniform margins.

Return curves - estimation

- We require a means of evaluating the joint tail, i.e., $\Pr(X > x, Y > y) = p$ for small p.
- Many bivariate extreme value models available for this purpose.
- Bivariate extreme value model = a framework for evaluating the dependence between extremes.
- How do we pick a model?

Return curves - estimation

- We restrict attention to two models due to their flexibility and minimal assumptions.
- In particular, neither make explicit assumptions about the form of extremal dependence.
- Model of Heffernan and Tawn (2004).
- Model of Wadsworth and Tawn (2013).

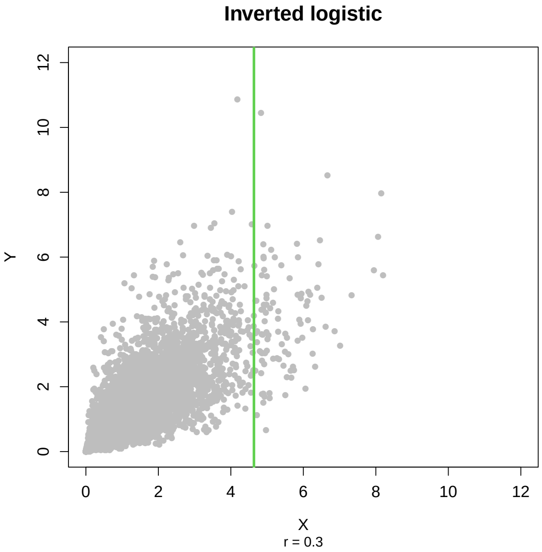
Return curves - estimation

- Suppose we have (X, Y) with standard exponential margins.
- Heffernan and Tawn (2004) assume that

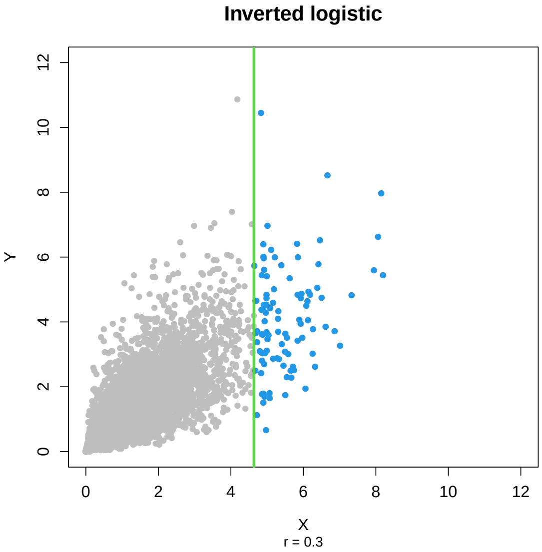
$$(Y \mid X = x) \sim \alpha x + x^{\beta} Z,$$

  for large x.
- $\alpha \in [0, 1]$, $\beta \in [0, 1)$ and Z is a residual process.
- α, β capture the extremal dependence of (X, Y).
- Fancy regression.

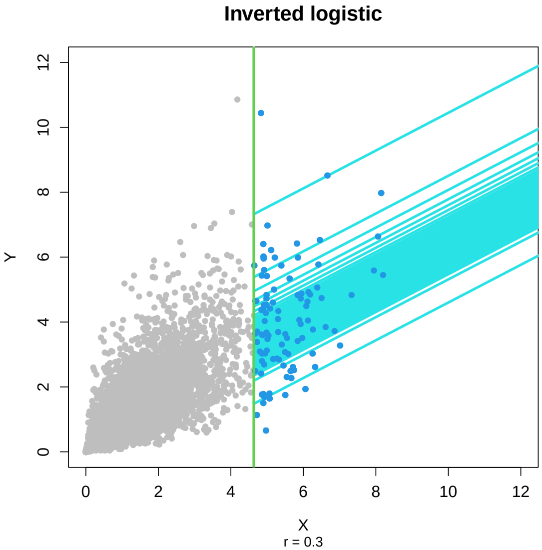
Return curves - estimation



Return curves - estimation



Return curves - estimation



Could also condition on large Y.

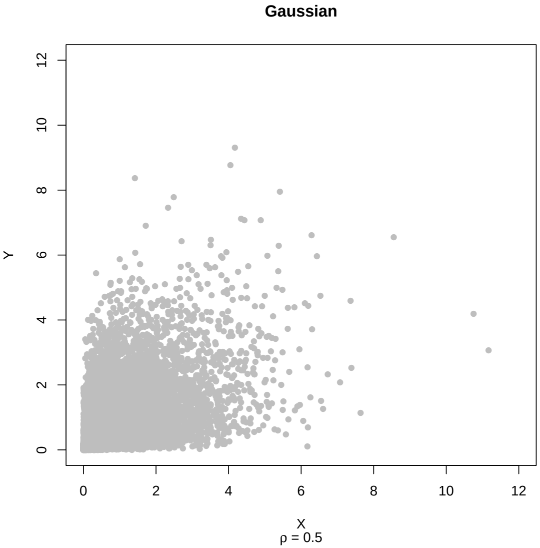
Return curves - estimation

- Suppose we again have (X, Y) on standard exponential margins.
- Let $K_w := \min\{X/w, Y/(1-w)\}$ for each $w \in [0, 1]$.
- Wadsworth and Tawn (2013) assume that for each $w \in [0, 1]$,

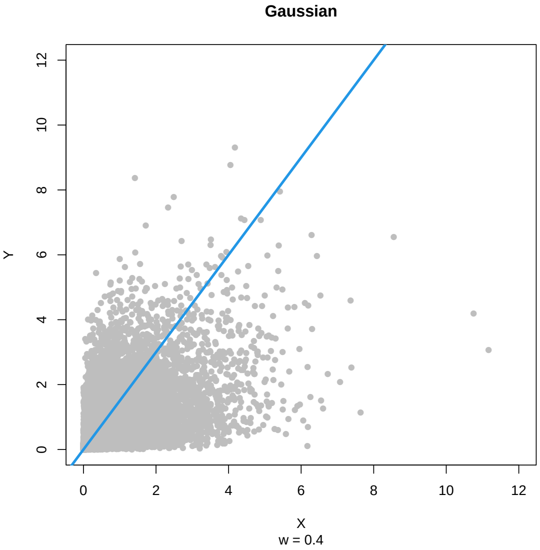
$$(K_w - u \mid K_w > u) \sim \text{Exp}(\lambda(w)),$$

  for large u.
- $\lambda(w)$ termed the angular dependence function.
- Captures the extremal dependence of (X, Y).

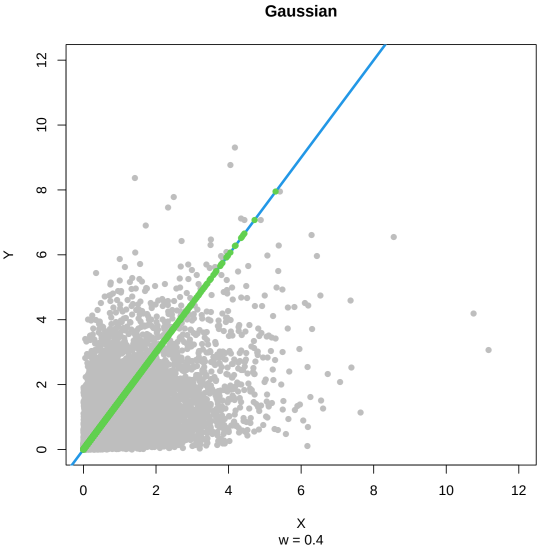
Return curves - estimation



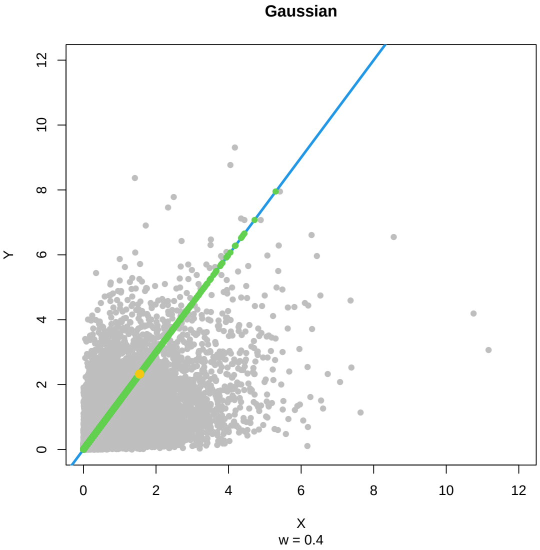
Return curves - estimation



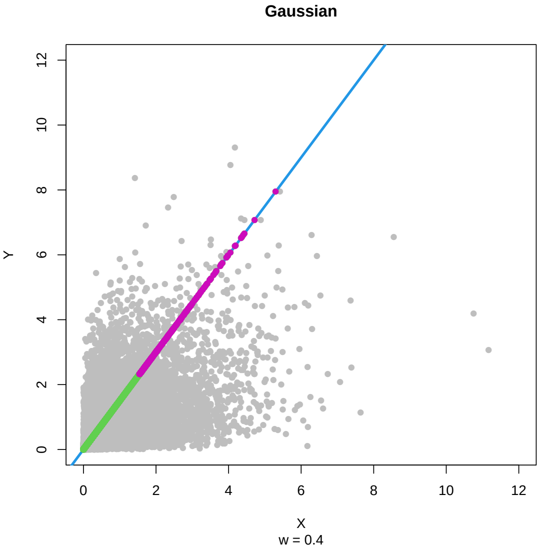
Return curves - estimation



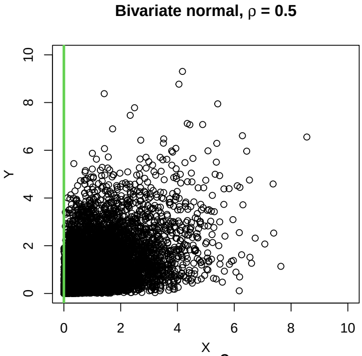
Return curves - estimation



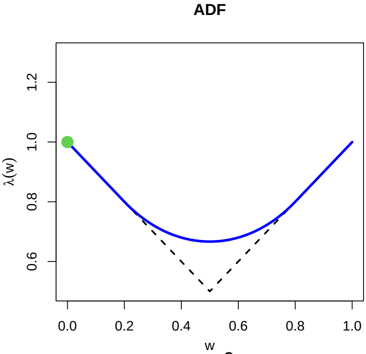
Return curves - estimation



Return curves - estimation



w = 0

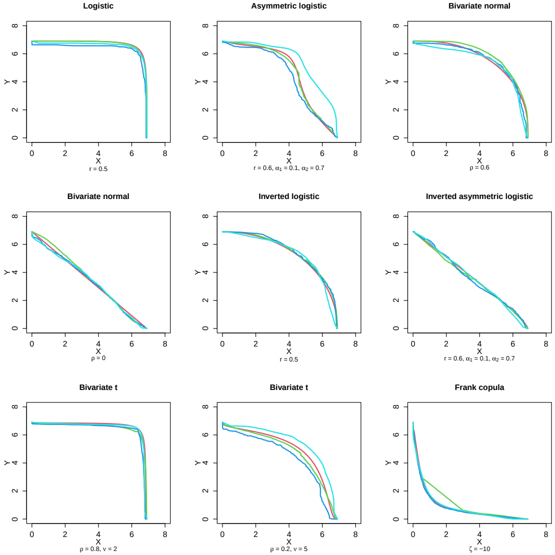


w = 0

Return curves - estimation

- Computed return curve estimates using Heffernan and Tawn (2004) and Wadsworth and Tawn (2013) models.
- Compared these estimates to an existing technique.
- Considered a range of copula examples.
- Goal: find the model with the least bias and variance.

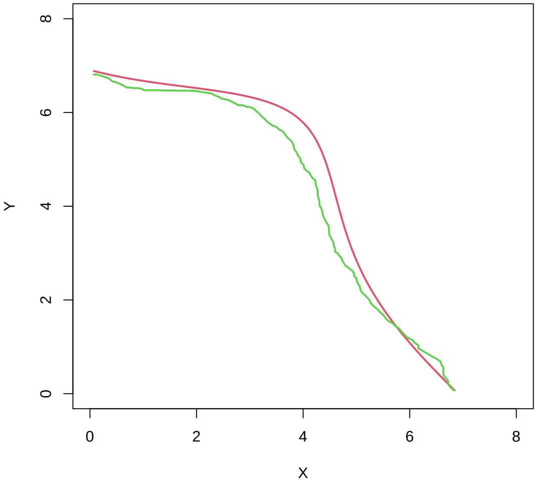
Return curves - estimation



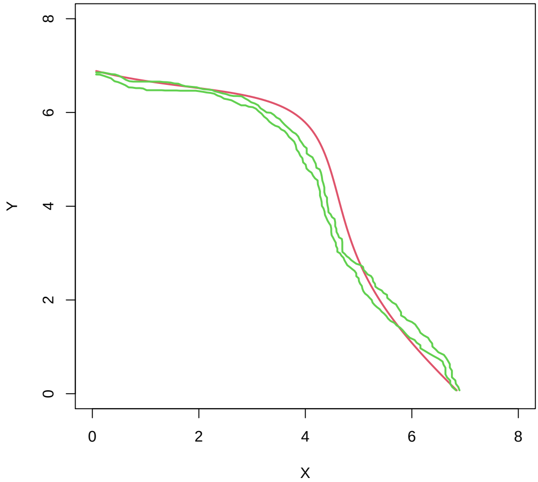
Return curves - estimation

- Model of Wadsworth and Tawn (2013) gave the best curve estimates overall.
- Quick, flexible, and outperformed the technique of Heffernan and Tawn (2004).
- This is significant since this latter approach is widely used for bivariate extreme value modelling.

Return curves - quantifying uncertainty



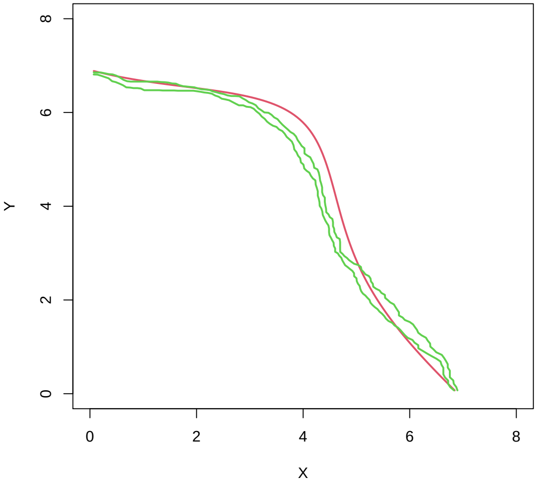
Return curves - quantifying uncertainty



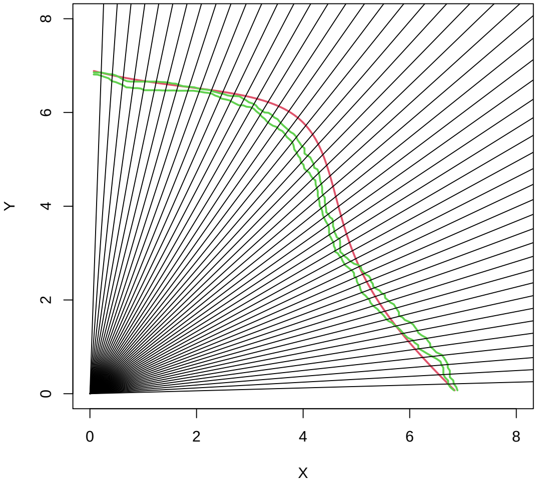
Return curves - quantifying uncertainty

- How do we capture variability in return curves?
- Estimates vary in two dimensions.

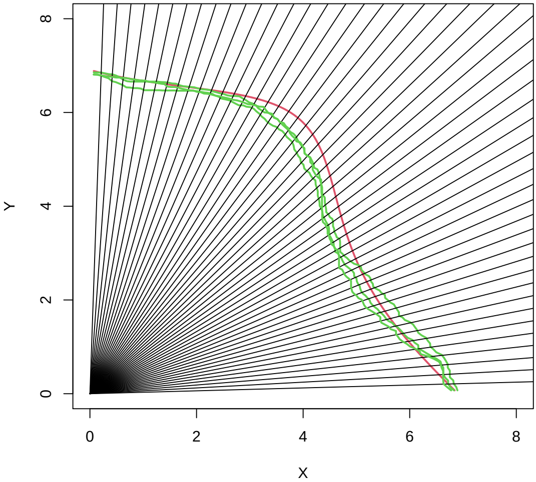
Return curves - quantifying uncertainty



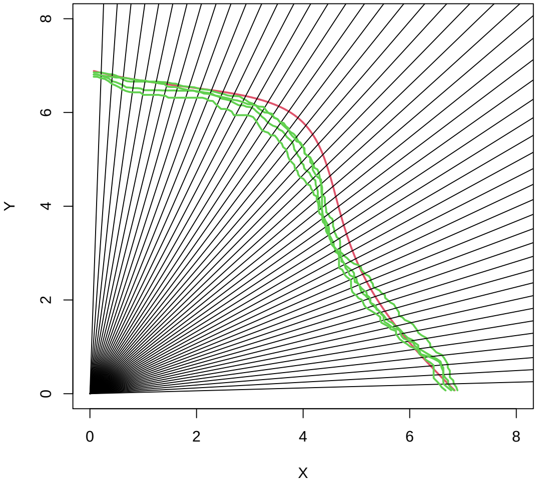
Return curves - quantifying uncertainty



Return curves - quantifying uncertainty



Return curves - quantifying uncertainty



Return curves - quantifying uncertainty

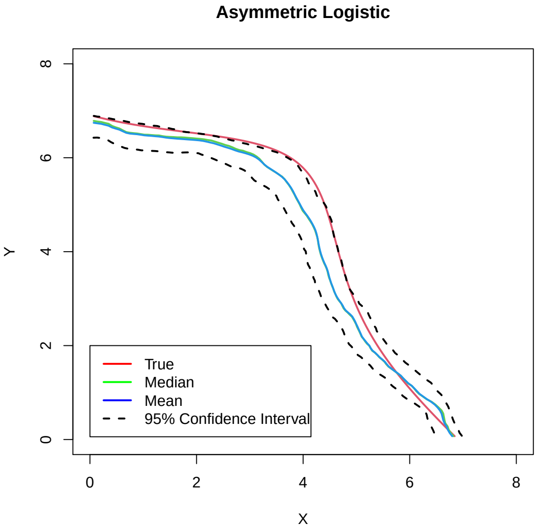
Formally:

- Define lines corresponding to a range of angles.
- Locate intersection points of curve estimates and lines.
- Compute Euclidean distances from origin to intersection points.
- Evaluate variability in distances.

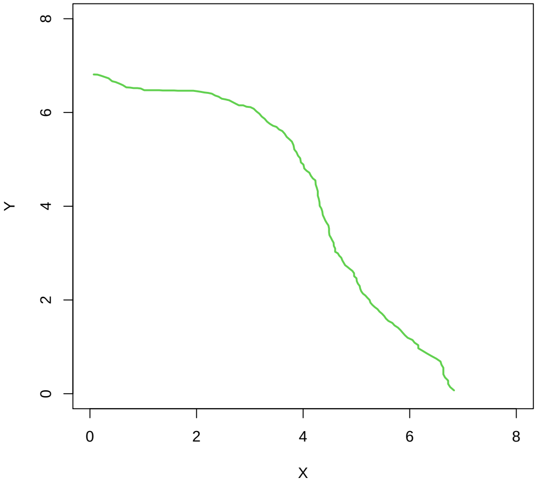
Return curves - quantifying uncertainty

- This reduces it to a one dimensional problem.
- Can compute meaningful confidence intervals.

Return curves - quantifying uncertainty



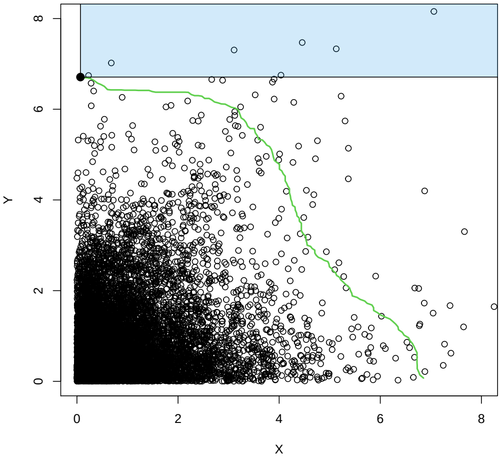
Return curves - diagnostics



Return curves - diagnostics

- In practical settings, we need to check goodness of fit for curve estimates.
- Does it accurately represent the probability p?
- ‘True’ curve is unknown.

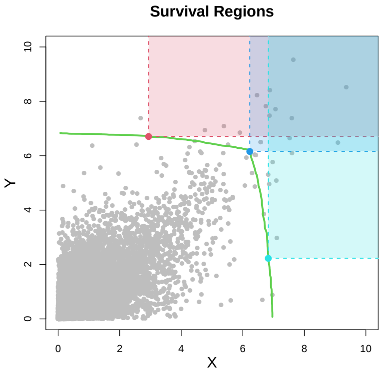
Return curves - diagnostics

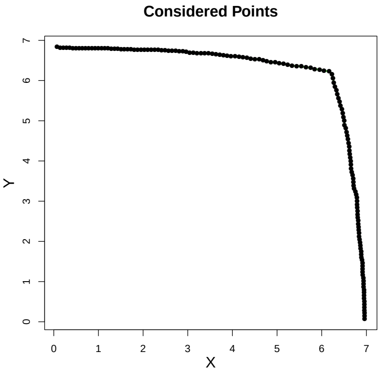


Return curves - diagnostics

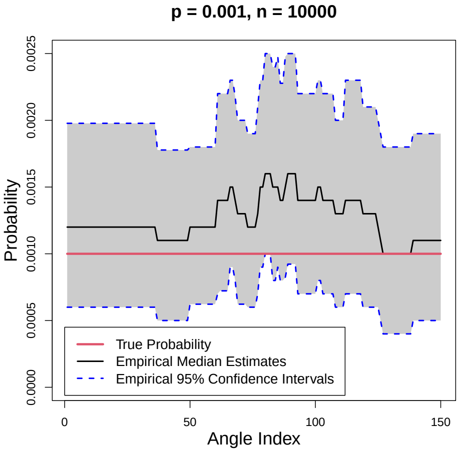
- Probability of being in any shaded region should equal p.
- We can estimate these probabilities empirically.
- Bootstrap the data to capture variability.
- Compare empirical estimates to true probability, which is pre-set.

Return curves - diagnostics





Return curves - diagnostics



Return curves

1. Return curve estimation ✓
2. Quantifying uncertainty ✓
3. Evaluating goodness of fit ✓

See Murphy-Barltrop et al. (2022) for further details.

Improved estimation of the angular dependence function

- Recall: Wadsworth and Tawn (2013) model gave the best return curve estimates overall.
- $\lambda(w)$ is the key model quantity.
- However, few applications of this model exist.
- Hence, estimation of λ not well studied.

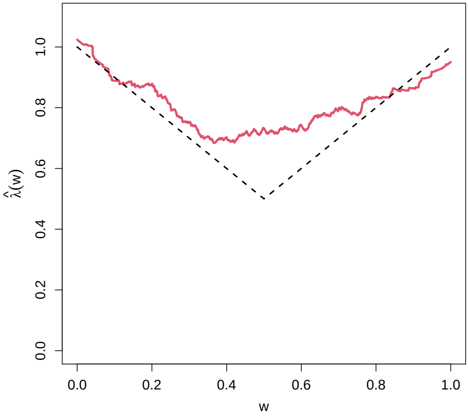
Improved estimation of the angular dependence function

- For each $w \in [0, 1]$ we assume,

$$(K_w - u \mid K_w > u) \sim \text{Exp}(\lambda(w)),$$

  for large u, where $K_w := \min\{X/w, Y/(1-w)\}$.
- Consequently, we can estimate $\lambda(w)$ using the maximum likelihood estimator for the exponential rate parameter.

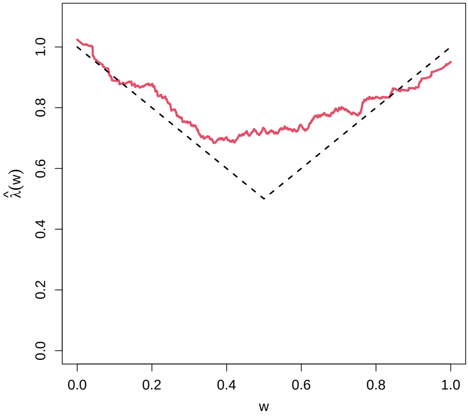
Improved estimation of the angular dependence function



Improved estimation of the angular dependence function

- This estimator is pointwise for each $w \in [0, 1]$.
- Bumpy and unrealistic.
- Does not respect theoretical conditions for angular dependence function:
  $\lambda(0) = \lambda(1) = 1$,
  $\lambda(w) \geq \max(w, 1 - w)$ for all $w \in [0, 1]$.

Improved estimation of the angular dependence function



Improved estimation of the angular dependence function

- We decided to address these issues.
- This required two steps:

1. Assuming a smooth, flexible form for λ.
2. Estimating parameters.

Improved estimation of the angular dependence function

- Bernstein-Bézier polynomials are a natural candidate.
- We assume

$$\lambda(w) = \sum_{i=0}^{k} \beta_i \binom{k}{i} w^i (1-w)^{k-i}, \ w \in [0,1],$$

for some $2 \leq k \in \mathbb{N}$, with $\beta_i \geq 0$ for all $i \leq k$.

Improved estimation of the angular dependence function

We want $\lambda(0) = \lambda(1) = 1$.

- Set $\beta_0 = \beta_k = 1$, giving

$$\lambda(w) = (1-w)^k + \sum_{i=1}^{k-1} \beta_i \binom{k}{i} w^i (1-w)^{k-i} + w^k, \; w \in [0,1],$$

Improved estimation of the angular dependence function

We want $\lambda(w) \geq \max(w, 1-w)$ for all $w \in [0,1]$.

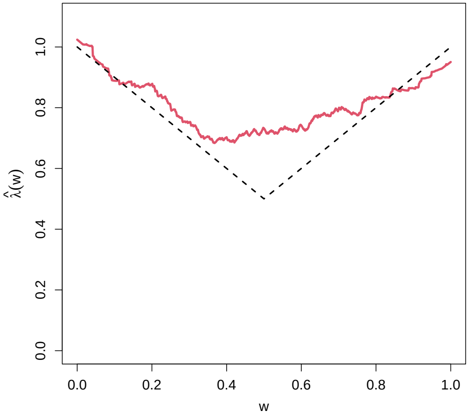
- Constrain parameter space of $(\beta_1, \ldots, \beta_{k-1})$ to satisfy this inequality.

$$\lambda(w) = (1-w)^k + \sum_{i=1}^{k-1} \beta_i \binom{k}{i} w^i (1-w)^{k-i} + w^k, \; w \in [0,1],$$

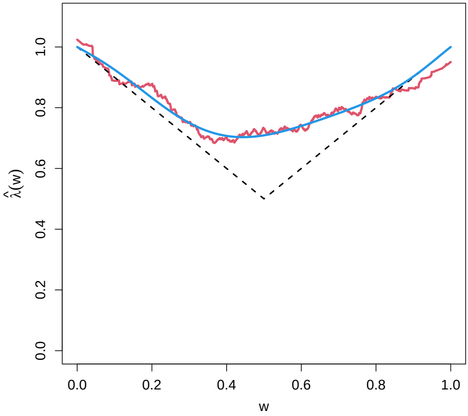
Improved estimation of the angular dependence function

- We use a composite likelihood approach to estimate the parameters $(\beta_1, \ldots, \beta_{k-1})$.
- Multiply likelihood functions together for different $w \in [0, 1]$.
- Better than considering each $w \in [0, 1]$ in turn.
- Gives a global estimator rather than a pointwise one.

Improved estimation of the angular dependence function

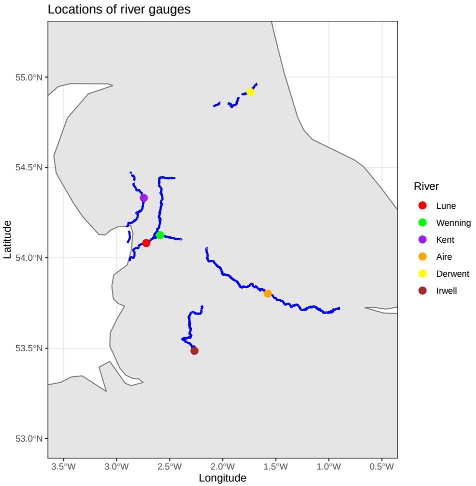


Improved estimation of the angular dependence function



Improved estimation of the angular dependence function

- We illustrate these techniques using river flow datasets from the North of England.



Improved estimation of the angular dependence function

- Understanding the probability of observing extreme river flow events (i.e., floods) at multiple sites simultaneously is important in various sectors, such as insurance (Keef et al., 2013) and environmental management (Lamb et al., 2010).
- How likely are we to observe joint floods for different pairs of sites?

Improved estimation of the angular dependence function

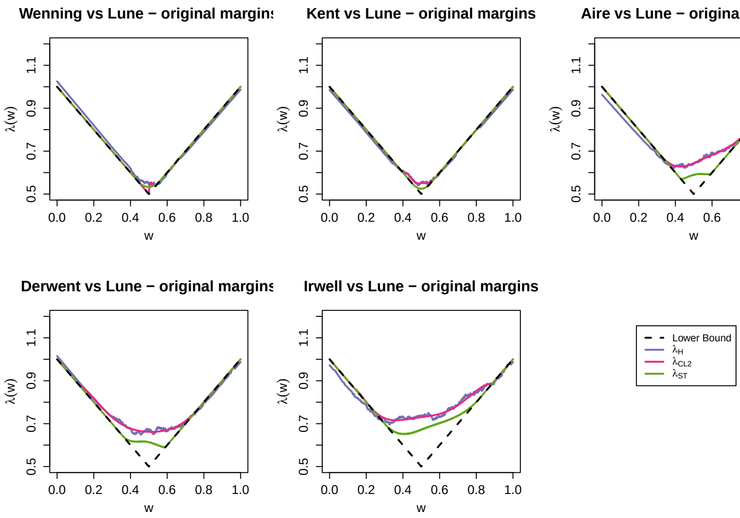
- Consider data over (approximately) 30 years for the October-March interval.
- Typically observe the highest rainfall in this period.
- Fix the site on the River Lune (near Lancaster) to be a reference site.

Improved estimation of the angular dependence function

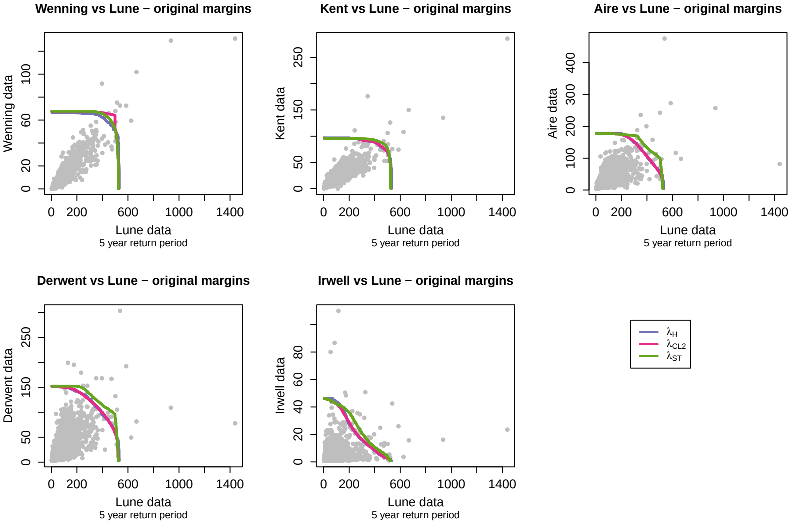
Steps

1. Transform marginal distributions to standard exponential.
2. Obtain estimates of the angular dependence function.
3. Estimate return curves for different pairs of gauges.

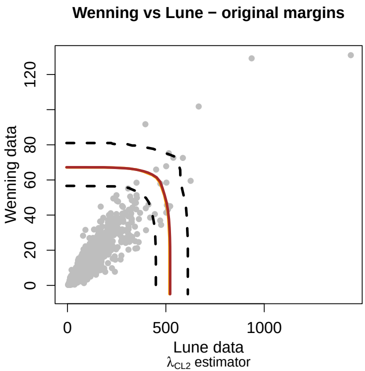
Improved estimation of the angular dependence function

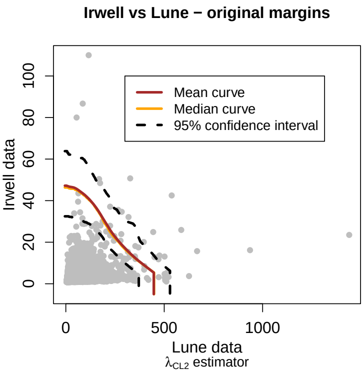


Improved estimation of the angular dependence function

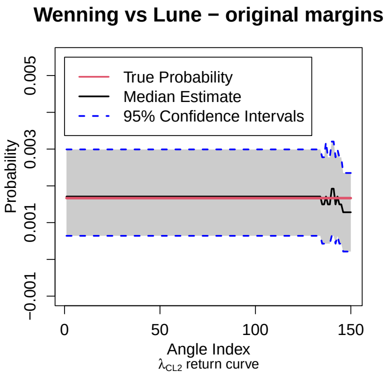


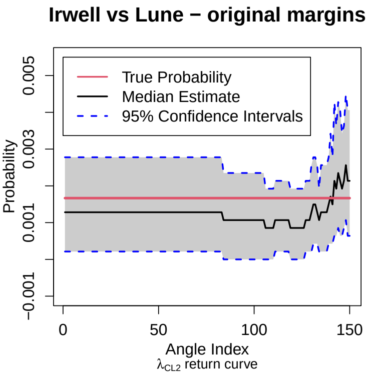
Improved estimation of the angular dependence function





Improved estimation of the angular dependence function





Improved estimation of the angular dependence function

- Novel estimates for the angular dependence function have lower bias and variance (on average)
- Allows us to better evaluate the extremal dependence structure.
- More accurate and realistic estimates of return curves.
- Preprint will be available very soon - keep an eye on arXiv.

Non-stationary return curves

- Up until now, we have assumed the data we are working with is independent and identically distributed.
- In practice, we often want to estimate return curves for combinations of environmental variables (e.g. temperature, relative humidity, wind speed).
- These variables can exhibit strong trends due to seasonality/climate change.
- Such variables are said to be non-stationary.

Non-stationary return curves

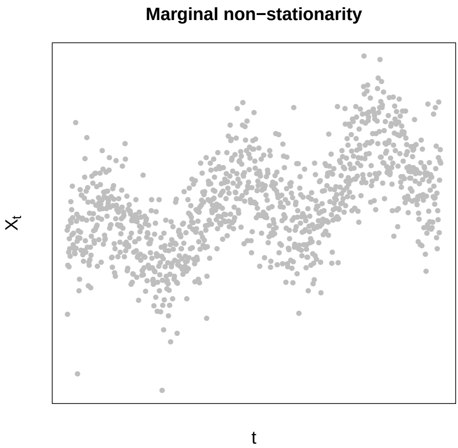
- Trends in variables = trends in return curves.
- Question: how do we model non-stationary trends in return curves?
- First need to consider what kinds of trend can exist.

Non-stationary return curves

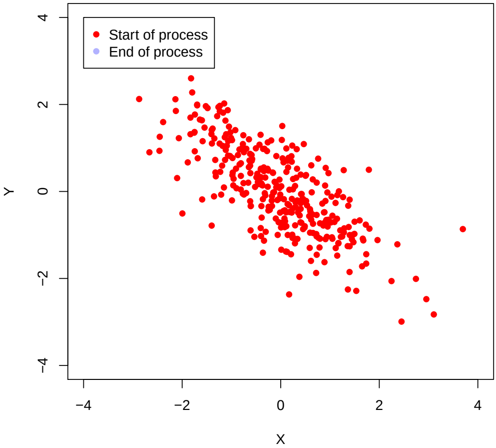
- Let $\{X_t, Y_t\}, t \in \{1, \ldots, n\}$ denote a bivariate process.
- Two possible forms of non-stationarity:

1. Trends in marginal distributions - individual behaviour of X_t and Y_t.
2. Trends in dependence structure - dependence between X_t and Y_t.

Non-stationary return curves



Non-stationary return curves



Non-stationary return curves

- Sparse literature for modelling non-stationarity in extremal dependence.
- Must account for both forms in our estimation procedure.
- Assume marginal non-stationarity has already been accounted for (i.e., marginal datasets have been de-trended).

Non-stationarity

We proposed an non-stationary extension of the Wadsworth and Tawn (2013) model.

- Let $\{X_t, Y_t\}$ denote a non-stationary process (on standard exponential margins) influenced by some external covariates $\mathbf{Z}_t, t \in \{1, \ldots, n\}$.
- For all $w \in [0, 1]$ and $t \in \{1, \ldots, n\}$, let $K_{w,t} := \min\left\{\frac{X_t}{w}, \frac{Y_t}{1-w}\right\}$.

Non-stationarity

- We assume

$$(K_{w,t} - u \mid K_{w,t} > u, \mathbf{Z}_t = \mathbf{z}_t) \sim \mathrm{Exp}(\lambda(w \mid \mathbf{Z}_t = \mathbf{z}_t))$$

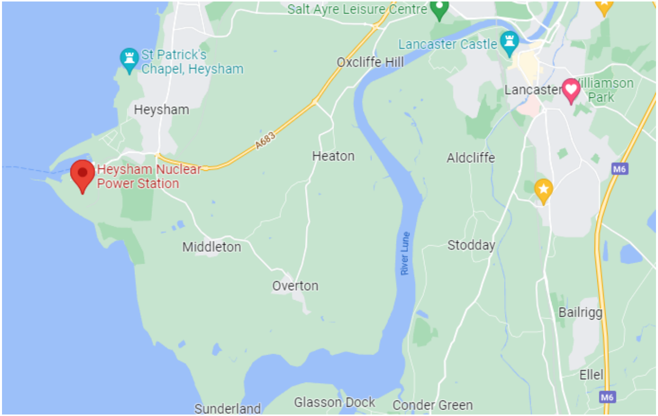
  for large u.
- $\lambda(w \mid \mathbf{Z}_t = \mathbf{z}_t)$ is termed the non-stationary angular dependence function.

Non-stationarity

- Can estimate $\lambda(w \mid \mathbf{Z}_t = \mathbf{z}_t)$ via standard quantile regression techniques.
- Flexible framework for capturing trends in the extremal dependence structure.
- Importantly, this model can be used to evaluate trends in return curves.
- Non-stationarity in return curve estimation has not been previously considered.

Non-stationary return curves

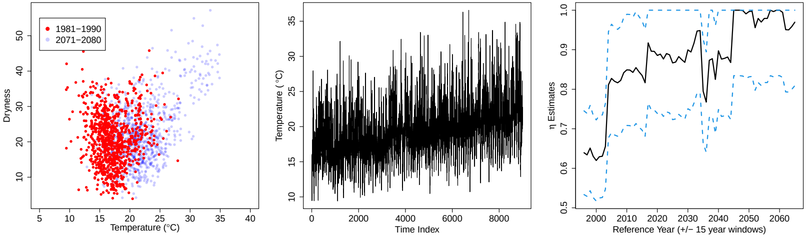
- Take the 1980-2080 temperature and relative humidity UKCP18 data for the nuclear site at Heysham, UK.



Non-stationary return curves

- We focus on summer data only.
- Let $0 < RH < 100$ represent relative humidity. We define a 'dryness' variable: $Dr := 100 - RH$.
- Combination of high temperature and high dryness typical of drought-like conditions.
- This is a concern for nuclear regulators (Knochenhauer and Louko, 2004).
- Understanding relationship between the extremes could allow for better risk management.

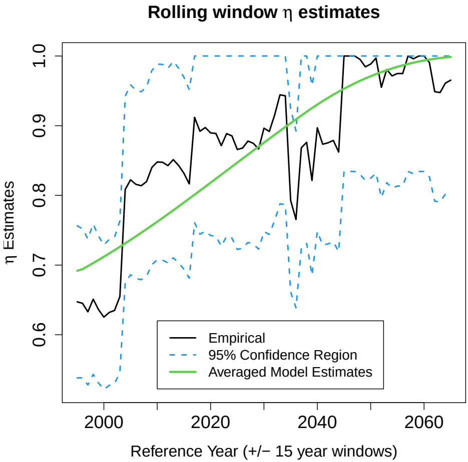
Non-stationary return curves



Non-stationary return curves

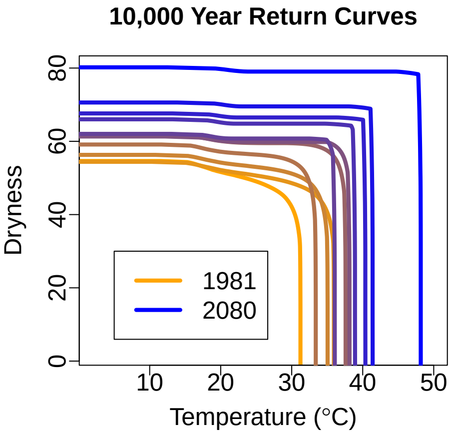
1. Remove marginal trends using methods proposed in Davison and Smith (1990) and Eastoe and Tawn (2009).
2. Transform data to standard exponential margins.
3. Fit non-stationary model.
4. Calculate return curve estimates up to the year 2080.

Non-stationary return curves



Non-stationary return curves

- Many nuclear facilities built to withstand 10^{-4} annual exceedance probability events.
- Such events will not be fixed in the non-stationary setting.



Non-stationary return curves

- Our model captures observed trends in extremal dependence.
- Can be used to estimate non-stationary return curves.
- Useful for evaluating joint extremal risks in future climates.
- See Murphy-Barltrop and Wadsworth (2022) for further details.

Discussion

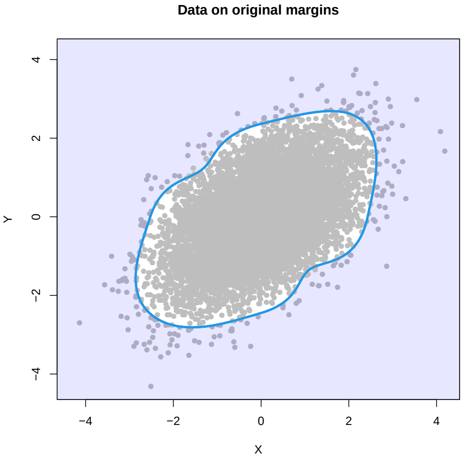
Still many open avenues within this research:

- Return curves in three (or more) dimensions.
- Further developing Wadsworth and Tawn (2013) modelling framework.
- Modelling negative extremal dependence structures.
- Alternative measures of risk.

... to name but a few.

Discussion

$$\Pr(\text{Outside Contour}) = p$$



Summary

- Motivation.
- Return curves.
- Improved estimation of the angular dependence function.
- Non-stationary return curves.
- Discussion.

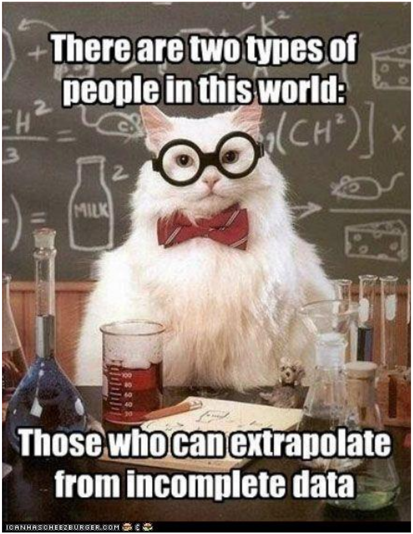
My thesis should be available through Lancaster University in June 2023 (approximately).

- This work will be used by the Office for Nuclear Regulation.
- Will help to quantify the risks from present and future joint extreme events.
- Could help inform the design bases for future nuclear installations.

Thanks very much for listening!

Does anyone have any questions? :)

c.barltrop@lancaster.ac.uk



References I

Coles, S. (2001). *An Introduction to Statistical Modeling of Extreme Values*. Springer London.

D'Arcy, E., Tawn, J. A., Joly, A., and Sifnioti, D. E. (2022). Accounting for seasonality in extreme sea level estimation. *arXiv*.

Davison, A. C. and Smith, R. L. (1990). Models for exceedances over high thresholds. *Journal of the Royal Statistical Society: Series B (Methodological)*, 52:393–425.

Eastoe, E. F. and Tawn, J. A. (2009). Modelling non-stationary extremes with application to surface level ozone. *Journal of the Royal Statistical Society. Series C: Applied Statistics*, 58:25–45.

Heffernan, J. E. and Tawn, J. A. (2004). A conditional approach for multivariate extreme values. *Journal of the Royal Statistical Society. Series B: Statistical Methodology*, 66:497–546.

Keef, C., Tawn, J. A., and Lamb, R. (2013). Estimating the probability of widespread flood events. *Environmetrics*, 24:13–21.

References II

Knochenhauer, M. and Louko, P. (2004). Guidance for external events analysis.

Lamb, R., Keef, C., Tawn, J., Laeger, S., Meadowcroft, I., Surendran, S., Dunning, P., and Batstone, C. (2010). A new method to assess the risk of local and widespread flooding on rivers and coasts. *Journal of Flood Risk Management*, 3:323–336.

Murphy-Barltrop, C. J. R. and Wadsworth, J. L. (2022). Modelling non-stationarity in asymptotically independent extremes. *arXiv*, 2203.05860.

Murphy-Barltrop, C. J. R., Wadsworth, J. L., and Eastoe, E. F. (2022). New estimation methods for extremal bivariate return curves. *arXiv*.

Wadsworth, J. L. and Tawn, J. A. (2013). A new representation for multivariate tail probabilities. *Bernoulli*, 19:2689–2714.