# Modelling non-stationarity in asymptotically independent extremes

C. J. R. Murphy-Barltrop[1,3] and J. L. Wadsworth[2]
[1]STOR-i Centre for Doctoral Training, Lancaster University LA1 4YR, United Kingdom
[2]Department of Mathematics and Statistics, Lancaster University LA1 4YF, United Kingdom
[3]Fathom, Square Works, 17-18 Berkeley Square, Clifton, Bristol BS8 1HB, United Kingdom

17th May 2022

STOR-i excellence with impact

Lancaster University

UKRI Engineering and Physical Sciences Research Council

ONR Office for Nuclear Regulation

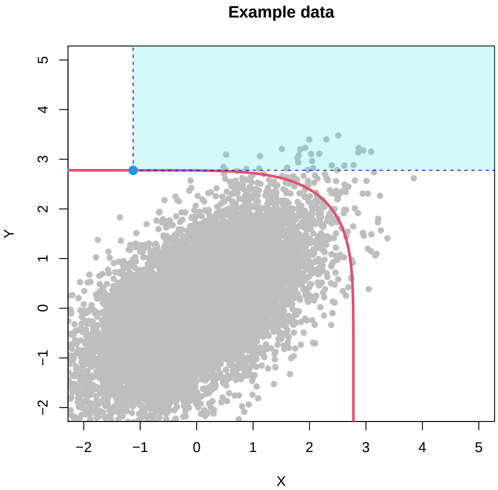
# Motivation - return curves

- My PhD: estimating **return curves** at extreme values.
- Given two variables,

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

  where $p$ is very small.
- Provide a summary of **extremal dependence**.

# Motivation - return curves



# Motivation - return curves

- In practice, we wish to estimate return curves for **combinations of environmental variables** (temperature, relative humidity, wind speed etc.).
- Such variables exhibit **non-stationarity**.

# Motivation - return curves

- Return curves **lack meaning** in the non-stationary setting, motivating an extended definition.
- Given $\{X_t, Y_t\}$ with **covariates $\mathbf{Z}_t$**, $t \in \{1, 2, \ldots, T\}$,

$$\mathrm{RC}_{\mathbf{z}_t}(p) := \{(x, y) \in \mathbb{R}^2 \mid \Pr(X_t > x, Y_t > y \mid \mathbf{Z}_t = \mathbf{z}_t) = p\}.$$

- **End goal**: estimating non-stationary return curves.

# Non-stationarity

- For $\{X_t, Y_t\}$, two forms of non-stationarity can exist.
  1. Trends in marginal extremes - tail of $X_t$ $(Y_t)$. **Well studied.**
  2. Trends in extremal dependence. **Very sparse literature.**
- Must account for both forms in our estimation procedure.
- We focus on second problem and present a new modelling technique.

# Multivariate extreme value theory (MVET)

- Marginal transformations: commonly transform marginal distributions prior to MVET analysis:
  1. $X_t \sim F_{X_t}$, $Y_t \sim F_{Y_t}$.
  2. $(U, V) := (F_{X_t}(X_t), F_{Y_t}(Y_t))$ has **UNIFORM** margins.
  3. Transform to margins of choice (standard exponential).
- Extremal dependence: commonly assess this feature via summary measures such as $\eta \in (0, 1]$ (Ledford and Tawn, 1996, 1997).
- $\eta = 1 \Rightarrow$ **asymptotic dependence**
- $\eta < 1 \Rightarrow$ **asymptotic independence**

$$\Pr(\min(X, Y) > u+v \mid \min(X, Y) > u) \to \exp\{-v/\eta\}, \quad u \to \infty$$

# Wadsworth and Tawn (2013) framework

Given $(X, Y)$ with standard exponential margins and **any ray** $w \in [0, 1]$:

1. Define $K_w := \min\left\{\frac{X}{w}, \frac{Y}{1-w}\right\}$.
2. Assume

$$\Pr\left\{K_w > u + v | K_w > u\right\} \to \exp\{-\lambda(w)v\},$$

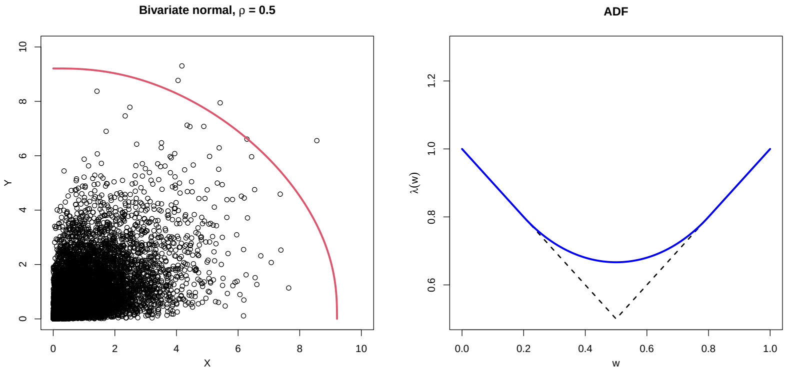
as $u \to \infty$ for any $v > 0$, with $\lambda(w) \geq \max(w, 1 - w)$.

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

# Wadsworth and Tawn (2013) framework

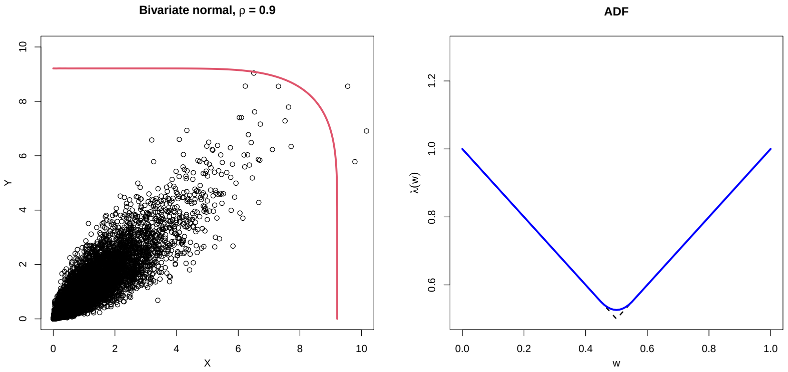
- $\lambda(w)$ is termed the **angular dependence function** (ADF) and determines the joint tail behaviour.
- Captures **both** asymptotic dependence and asymptotic independence - unlike many MVET approaches.
- **Extension** of Ledford and Tawn (1996, 1997), with $\eta = 1/(2\lambda(0.5))$.
- Allows evaluation of extremal dependence in all regions (see next slide).

# Examples

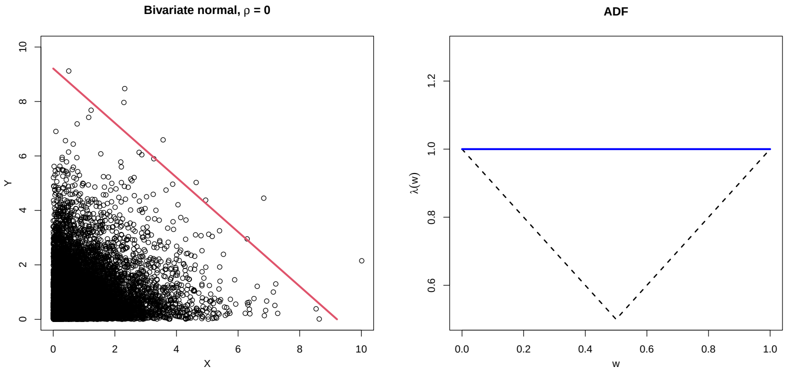


Bivariate normal, ρ = 0.5
Y
X
0
2
4
6
8
10
ADF
λ(w)
w
0.6
0.8
1.0
1.2
0.0
0.2
0.4
0.6
0.8
1.0

# Examples

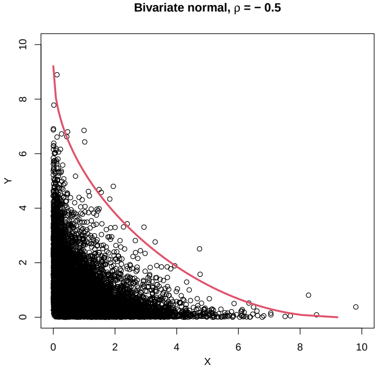

Bivariate normal, ρ = 0.9
Y
X
0
2
4
6
8
10
ADF
λ(w)
w
0.6
0.8
1.0
1.2
0.0
0.2
0.4
0.6
0.8
1.0

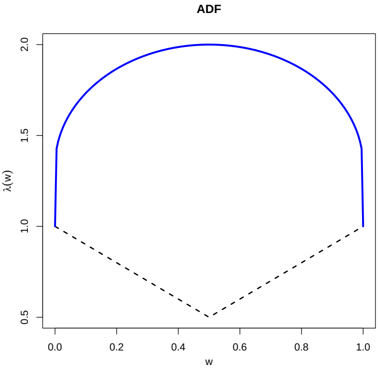
# Examples



Bivariate normal, ρ = 0
Y
X
0
2
4
6
8
10
ADF
λ(w)
w
0.6
0.8
1.0
1.2
0.0
0.2
0.4
0.6
0.8
1.0

# Examples


Bivariate normal, ρ = − 0.5
X
Y
0
2
4
6
8
10


ADF
w
λ(w)
0.0
0.2
0.4
0.6
0.8
1.0
0.5
1.0
1.5
2.0

# Non-stationary extension

$\{X_t, Y_t\}$ with standard exponential margins and any ray $w \in [0, 1]$.

1. Define $K_{w,t} := \min\left\{\frac{X_t}{w}, \frac{Y_t}{1-w}\right\}$.
2. Assume

$$\Pr\left(K_{w,t} > v + u \middle| K_{w,t} > u, \mathbf{Z}_t = \mathbf{z}_t\right) \to \exp\{-\lambda(w \mid \mathbf{Z}_t = \mathbf{z}_t)v\},$$

as $u \to \infty$ for any $v > 0$ and $t \leq T$.

$\lambda(w \mid \mathbf{Z}_t = \mathbf{z}_t)$ is termed the **non-stationary** ADF.

# Estimation

- Select high quantiles $q_1 < q_2$ close to 1.
- For a **fixed** $w$, we use quantile regression to estimate $u_1$ and $u_2$ such that

$$\Pr\left(K_{w,t} \leq u_1 \middle| \mathbf{Z}_t = \mathbf{z}_t\right) = q_1$$
$$\Pr\left(K_{w,t} \leq u_2 \middle| \mathbf{Z}_t = \mathbf{z}_t\right) = q_2.$$

- As $q_1 \to 1$, we have

$$\frac{1-q_2}{1-q_1} = \exp\{-\lambda(w \mid \mathbf{Z}_t = \mathbf{z}_t)(u_2 - u_1)\}$$

# Estimation

- Estimator given by

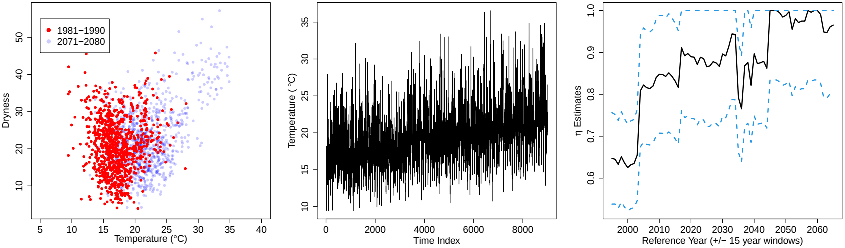
$$\hat{\lambda}(w \mid \mathbf{Z}_t = \mathbf{z}_t) = -\frac{1}{u_2 - u_1} \log\left(\frac{1 - q_2}{1 - q_1}\right)$$

- In practice, we average over quantile levels and smooth the estimator over $w$ using Bernstein polynomials (Marcon et al., 2017).
- See Murphy-Barltrop and Wadsworth (2022) for further details.

# Motivating example - UKCP18 data

- Take the 1980-2080 **temperature** and **relative humidity** projections for the nuclear site at Heysham, UK - we focus on summer data only.
- Suppose $0 < RH < 100$ represents relative humidity. We define a **'dryness'** variable: $Dr := 100 - RH$.
- **Combination** of high temperature and high dryness relevant for **nuclear safety** (Knochenhauer and Louko, 2004).
- Understanding relationship between the extremes could allow for better risk management.

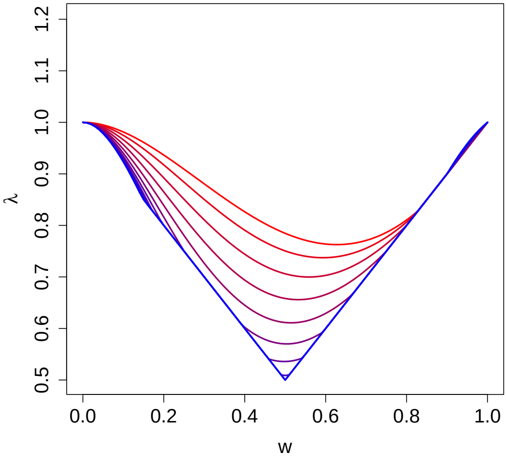
# Trends in the data



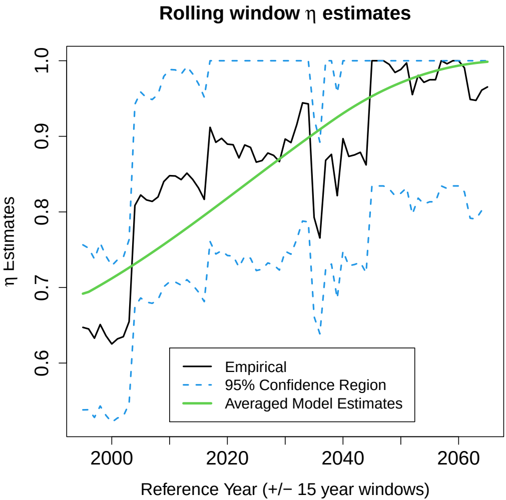
# Modelling procedure

1. **Remove marginal trends** using methods proposed in Davison and Smith (1990) and Eastoe and Tawn (2009).
2. **Transform** data to **exponential margins**.
3. Obtain **estimate** of **non-stationary ADF** through quantile regression.
4. **Calculate** return curve estimates up to the **year 2080**.

# Non-stationary ADF



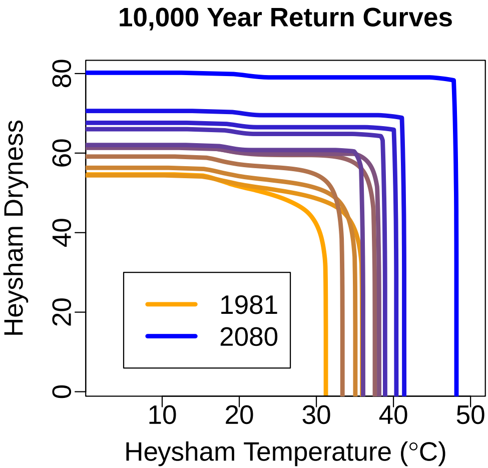
# $\eta$ estimates




(Model estimate given by $\hat{\eta}_t = [2\hat{\lambda}(0.5 \mid \mathbf{Z}_t = \mathbf{z}_t)]^{-1}$ )

# Summer 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.



# Conclusions

- We are able to estimate non-stationary return curves that **reflect observed trends**.
- Our work makes a contribution to a particularly sparse field.

**Thanks for listening! Does anyone have any questions?**

# References I

Davison, A. C. and Smith, R. L. (1990). Models for Exceedances Over High Thresholds. *Journal of the Royal Statistical Society: Series B (Methodological)*, 52(3):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(1):25–45.

Knochenhauer, M. and Louko, P. (2004). Guidance for External Events Analysis. In *Probabilistic Safety Assessment and Management*, pages 1498–1503. Springer, London.

Ledford, A. W. and Tawn, J. A. (1996). Statistics for near independence in multivariate extreme values. *Biometrika*, 83(1):169–187.

Ledford, A. W. and Tawn, J. A. (1997). Modelling dependence within joint tail regions. *Journal of the Royal Statistical Society. Series B: Statistical Methodology*, 59(2):475–499.

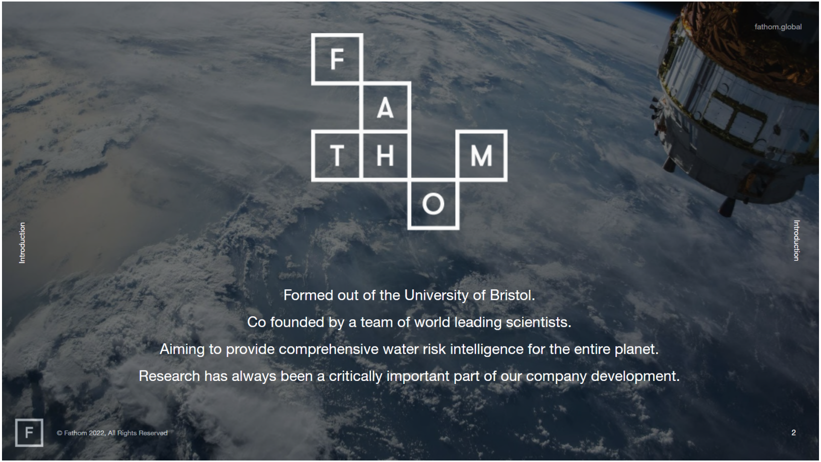
# References II

Marcon, G., Padoan, S. A., Naveau, P., Muliere, P., and Segers, J. (2017). Multivariate nonparametric estimation of the Pickands dependence function using Bernstein polynomials. *Journal of Statistical Planning and Inference*, 183:1–17.

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

Wadsworth, J. L. and Tawn, J. A. (2013). A new representation for multivariate tail probabilities. *Bernoulli*, 19(5 B):2689–2714.

# One final note



## First publication of large scale flood dependency in the US

Flood spatial dependency is defined using the observed record.

Fathom's event set covers 10,000 years of simulation and >500,000 unique events.

Water Resources Research

Research Article | Open Access

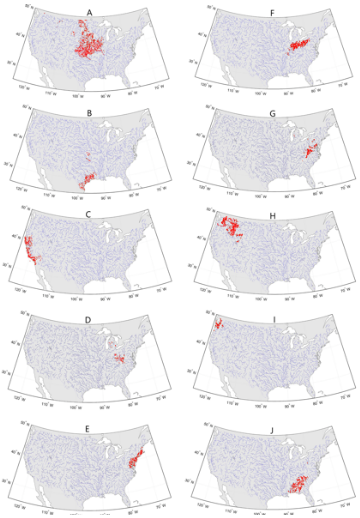
**The Spatial Dependence of Flood Hazard and Risk in the United States**

Niall Quinn, Paul D. Bates, Jeff Neal, Andy Smith, Oliver Wing, Chris Sampson, James Smith, Janet Heffernan

First published: 13 February 2019 | https://doi.org/10.1029/2018WR024205 | Citations: 14

SECTIONS PDF TOOLS SHARE

Abstract




Please come and speak to me at any point if you have any questions/queries about the research we do!

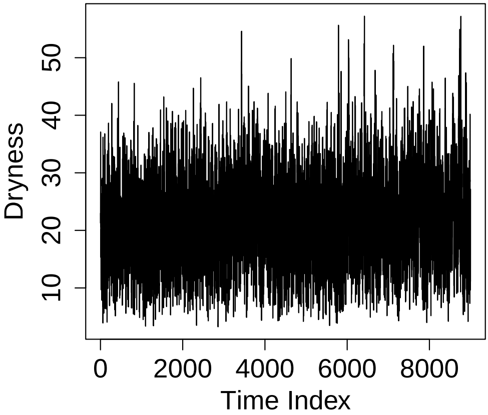
# Averaging over quantiles

- Select some high quantiles $q_{1,i} < q_{2,i}$ close to 1 with $i \leq n \in \mathbb{N}$.
- Use quantile regression to estimate $u_{1,i}$ and $u_{2,i}$ such that

$$\Pr\left(K_{w,t} \leq u_{1,i} \middle| \mathbf{Z}_t = \mathbf{z}_t\right) = q_{1,i}$$
$$\Pr\left(K_{w,t} \leq u_{2,i} \middle| \mathbf{Z}_t = \mathbf{z}_t\right) = q_{2,i}.$$

- $$\lambda(w \mid \mathbf{Z}_t = \mathbf{z}_t) \approx -\frac{1}{n}\sum_{i=1}^{n} \frac{1}{u_{2,i} - u_{1,i}} \log\left(\frac{1 - q_{2,i}}{1 - q_{1,i}}\right)$$

# Dryness



# Marginal trends