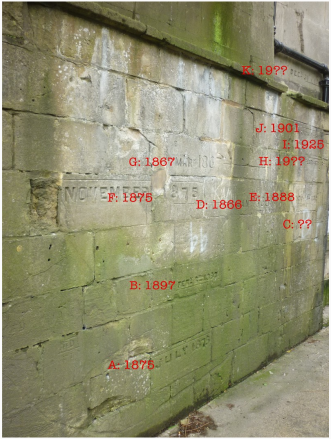
A clustering framework for conditional extremes models

SAMBa Conference 2025 Talk,
11th July, 2025,

Paddy O'Toole, University of Bath
Supervised by Christian Rohrbeck and Jordan Richards (University of Edinburgh)

Introduction


K: 19??
J: 1901
I: 1925
G: 1867
H: 19??
F: 1875
D: 1866
E: 1888
C: ??
B: 1897
A: 1875

Problem

- Often want to estimate

$$\mathbb{P}(X > x, Y > y) = \mathbb{P}(Y > y \mid X > x)\mathbb{P}(X > x)$$

  for large x, y
- "concomitant"/concurrent extreme events for random vector $\mathbf{X}$ often particularly devastating
- Goal: **identify trends by clustering sites with similar tail dependence**

Dependence Modelling

Asymptotic dependence

- Coefficient of extremal dependence $\chi \in [0, 1]$,

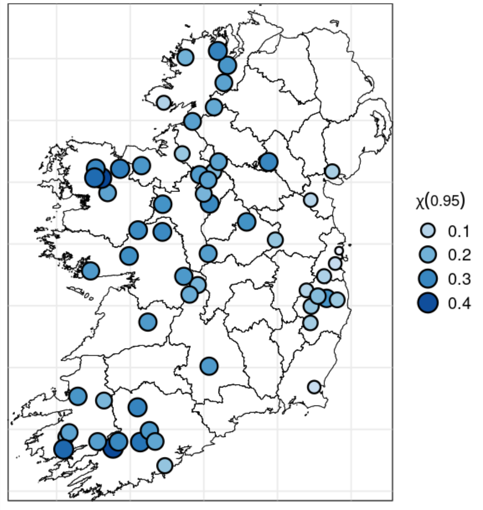
$$\chi = \lim_{u \to 1} \mathbb{P}\big[F_1(X_1) > u \mid F_2(X_2) > u\big]$$

- (Increasingly strong) asymptotic dependence for $\chi > 0$.
- However, χ only gives summary; inference requires **dependence model**.

Coles, J. Heffernan, and J. Tawn, *“Dependence measures for extreme value analyses”* (1999)

Ireland

- Precipitation[1] & wind speed[2] data for 59 sites across Ireland, Winter months (Oct-Mar) 1990-2020

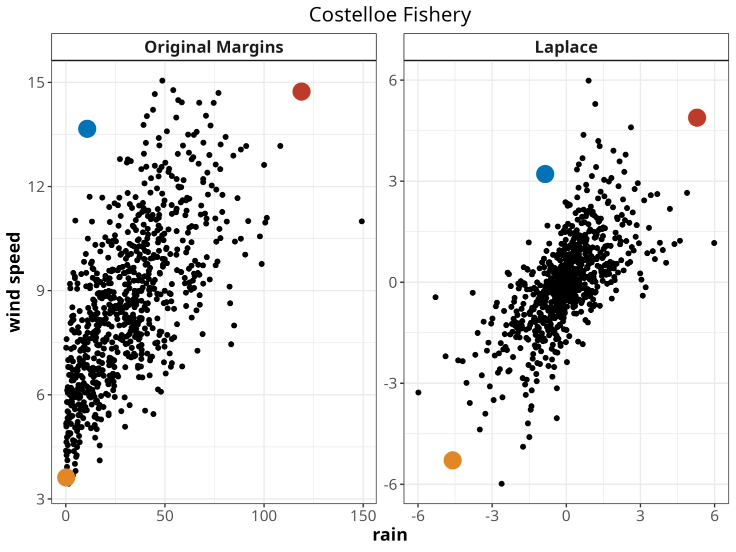


[1] Met Éireann weekly aggregate

[2] ERA5 reanalysis weekly mean of daily maxima

Conditional extremes

Marginal transformation



Conditional extremes

- Heteroskedastic regression dependence model:

$$\big(Y \mid X = x\big) = \alpha x + x^{\beta} Z, \text{ for } x > u$$

  - slope parameter α for Y given large X,
  - "spread" parameter $\beta \in (-\infty, 1]$ controls stochasticity of relationship between Y and large X.

Conditional extremes

- Heteroskedastic regression dependence model:

$$\left(\boldsymbol{Y}_{-i} \mid Y_i = y_i\right) = \boldsymbol{\alpha}_{|i}\boldsymbol{y}_i + \boldsymbol{y}_i^{\boldsymbol{\beta}_{|i}}\boldsymbol{Z}_{|i}, \text{ for } y_i > u_{Y_i}$$

  - slope parameter $\alpha_{j|i} \in [-1, 1]$ for Y_j given large Y_i,
  - "spread" parameter $\beta_{j|i} \in (-\infty, 1]$ controls stochasticity of relationship between Y_j and large Y_i.

J. E. Heffernan and J. A. Tawn, *"A Conditional Approach for Multivariate Extreme Values (with Discussion)"* (2004)

Conditional extremes

- **Heteroskedastic regression dependence model**:

$$\left(\boldsymbol{Y}_{-i} \mid Y_i = y_i\right) = \boldsymbol{\alpha}_{|i}\boldsymbol{y}_i + \boldsymbol{y}_i^{\boldsymbol{\beta}_{|i}}\boldsymbol{Z}_{|i}, \text{ for } y_i > u_{Y_i}$$

  - slope parameter $\alpha_{j|i} \in [-1, 1]$ for Y_j given large Y_i,
  - "spread" parameter $\beta_{j|i} \in (-\infty, 1]$ controls stochasticity of relationship between Y_j and large Y_i.

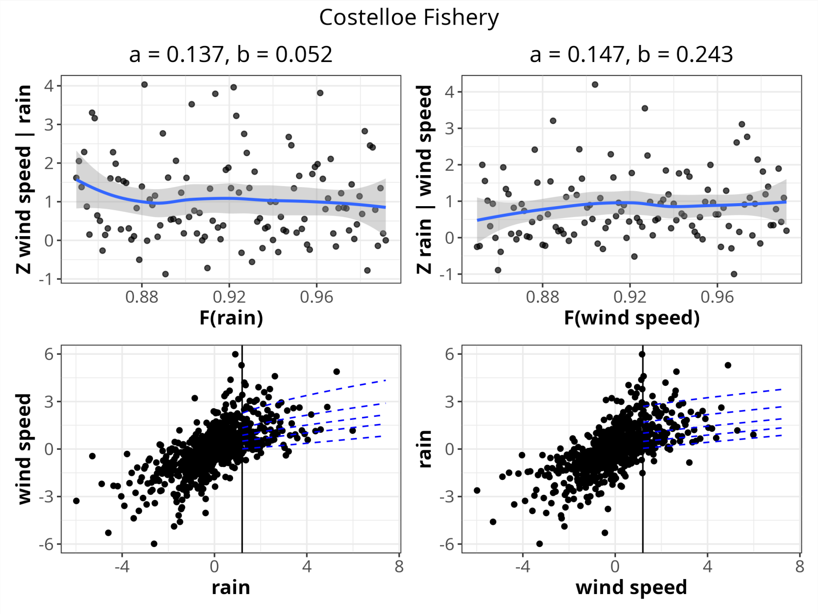
- Key assumptions:
  - Residuals $\boldsymbol{Z}_{|i} \sim N(\boldsymbol{\mu}_{|i}, \boldsymbol{\Sigma}_{|i})$
  - $\boldsymbol{Z}_{|i}, Y_i$ conditionally independent for large Y_i

- Special cases:

$$\begin{aligned}
\boldsymbol{\alpha}_{|i} = 0, \boldsymbol{\beta}_{|i} = 0 &\implies \boldsymbol{Y}_{-i}, \boldsymbol{Y}_i \text{ independent,} \\
\boldsymbol{\alpha}_{|i} = -1/1, \boldsymbol{\beta}_{|i} = 0 &\implies \text{perfect positive/negative dependence,} \\
-1 < \boldsymbol{\alpha}_{|i} < 1 &\implies \text{asymptotic independence}
\end{aligned}$$

J. E. Heffernan and J. A. Tawn, *"A Conditional Approach for Multivariate Extreme Values (with Discussion)"* (2004)

Conditional extremes



Inference

Inference assumes conditional distribution follows a multivariate Normal (MVN) distribution:

$$\left(\boldsymbol{Y}_{-i} \mid Y_i = y_i\right) \sim N\left(\boldsymbol{\alpha}_{|i} y_i + y_i^{\boldsymbol{\beta}_i} \boldsymbol{\mu}_{|i}, y_i^{\boldsymbol{\beta}_i} \boldsymbol{\Sigma}_{|i}\right), \text{ for } Y_i > u_{Y_i}$$

$\implies$ **dependence structures at different sites can be compared using their MVN distributions**

Clustering