Investigating the effect of aggregation on extremal dependence

Dylan Bahia[1] Jordan Richards[2]

[1]University of Manchester; [2]STOR-i, Lancaster University

STOR-i excellence with impact

Lancaster University

Introduction

- There are two ways of classifying an extreme event.
- The first method is to split the time series into fixed blocks, and take the maximum of each block to be extreme:
  - Let X_i denote the i^{th} observation of a random variable X.
  - Let $M_n = max\{X_1, ..., X_n\}$.
  - Let M_n^* be the normalised maxima, satisfying

$$P(M_n^* \leq z) \to G(z) \text{ as } n \to \infty,$$

  where G is a non-degenerate distribution function.
- Then G is a Generalised Extreme Value (GEV) distribution, where

$$G(z) = exp\left(-\left(1 + \xi\left({}^{(z-\mu)}/_{\sigma}\right)\right)^{-1/\xi}\right)$$

- There are three parameters:
  - The location parameter μ, where $-\infty < \mu < \infty$
  - The scale parameter σ, where $\sigma > 0$
  - The shape parameter ξ, where $-\infty < \xi < \infty$
- This distribution is defined on the set

$$\{z : 1 + {}^{\xi(z-\mu)}/_{\sigma} > 0\}$$

- The second method of classifying extremes is to choose a threshold and take all values above this threshold to be extreme:
  - Assume that the block maxima are GEV distributed.
- Then for large enough u, the distribution of $(X - u)$ given $X > u$ is approximately a Generalised Pareto Distribution (GPD), defined by

$$H(y) = P(X > u + y \mid X > u) \\ = 1 - (1 + {}^{\xi y}/_{\tilde{\sigma}})^{-1/\xi}$$

  where $\tilde{\sigma} = \sigma + \xi(u - \mu)$
- The distribution is defined on the set

$$\{y : y > 0 \ \& \ (1 + {}^{\xi y}/_{\tilde{\sigma}}) > 0\}$$

- The ξ of both the GEV and GPD are identical given the same data set.

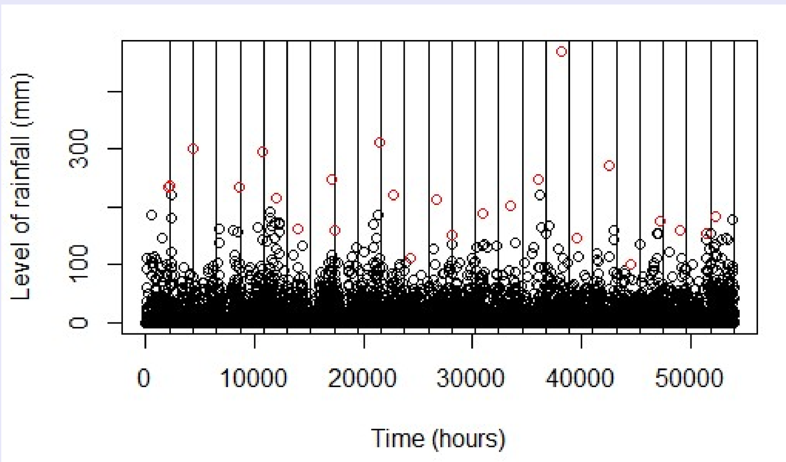

Figure 1: Local maxima of rainfall at fixed location over time

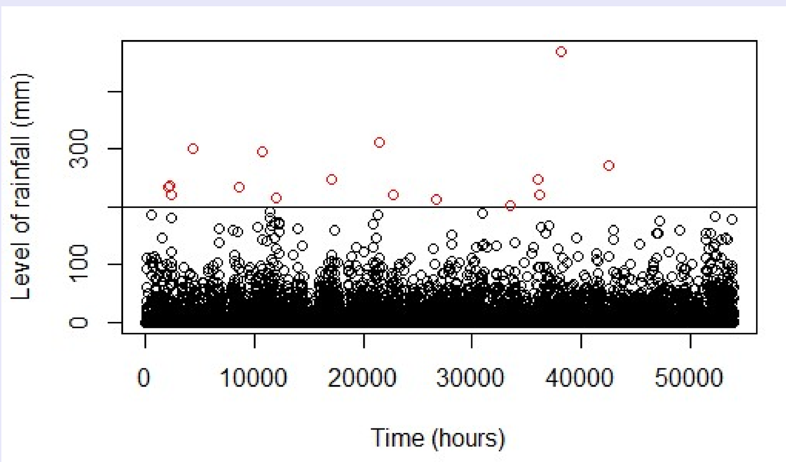

Figure 2: Rainfall observations above threshold at fixed location over time

Return levels

- A series of independent observations X_i, $i = 1, ..., d$, can be blocked into sequences of length n (the length is often 1 year), generating a series of block maxima $M_{n,1}, ...M_{n,m}$, where m is the number of maxima.
- A GEV distribution can be fitted to these maxima, and extreme quantiles can be estimated by inverting the distribution function:

$$z_p = \begin{cases} \mu - \frac{\sigma}{\xi}(1 - (-log(1-p))^{-\xi}) & \text{for } \xi \neq 0 \\ \mu - \sigma \log(-\log(1-p)) & \text{for } \xi = 0 \end{cases}$$

  where $G(z_p) = 1 - p$.
- z_p is the return level associated with the return period $1/p$, i.e. z_p is expected to be exceeded every $1/p$ years.

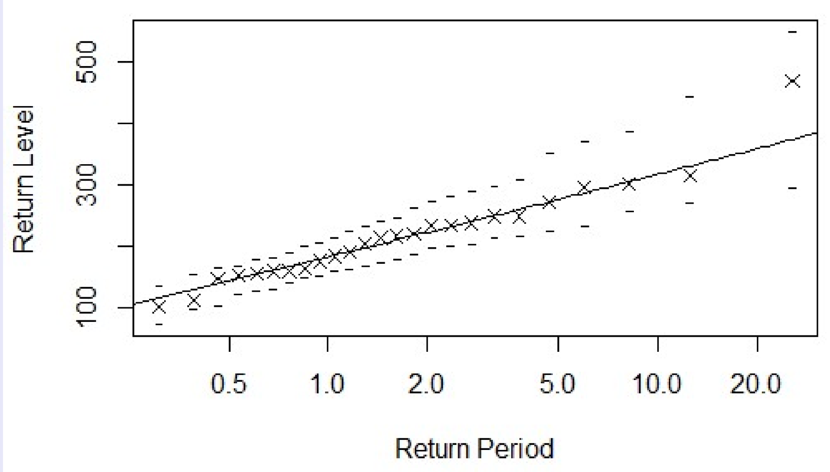

Figure 3: Expected return level after given period of years

Extremal Dependence Measures

- Asymptotic dependence is a measure of extremes occurring together.
- One dependence measure is χ, which is the upper tail index. Given two random variables X and Y, and their respective uniform transformations U and V,

$$\chi = lim_{u \to 1} P(V > u \mid U > u)$$

- This can be calculated empirically.
- χ takes values between 0 and 1, with 1 signifying perfect asymptotic dependence and 0 signifying asymptotic independence.
- The data below has high asymptotic dependence.

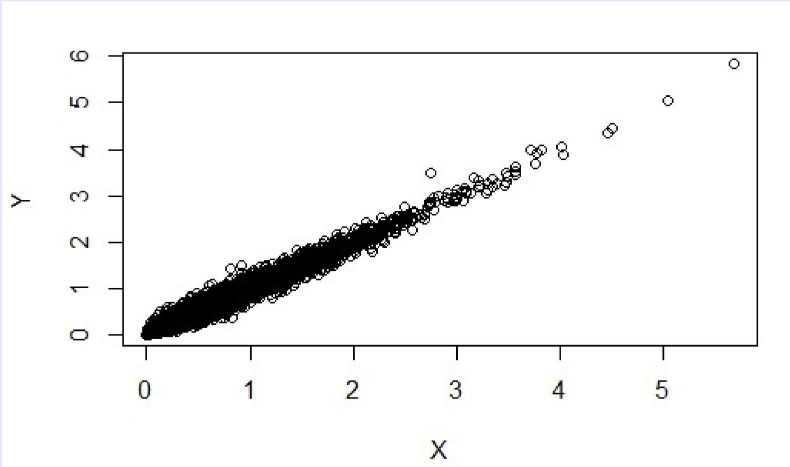

Figure 4: Simulation of random variables X and Y such that $\chi = 0.9$

Extremal Dependence Measures (cont.)

- Another dependence measure is η, which is the coefficient of tail dependence. Given two random variables X and Y on exponential margins,

$$P(X > x, Y > x) \sim L(x)exp(-{}^{x}/_{\eta})$$

  where $L(x)$ is a slowly varying function
- It can be estimated by Hill's estimator, and takes values between 0.5 and 1, with 1 signifying asymptotic dependence.
- The data below has low asymptotic dependence

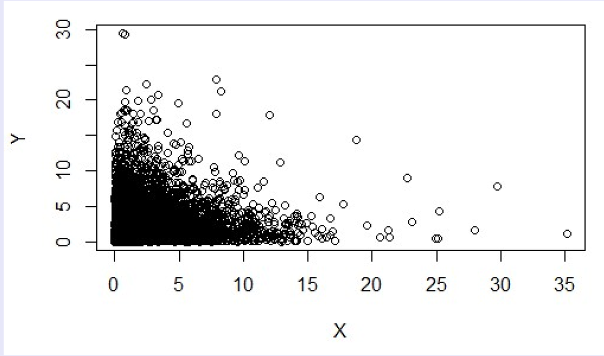

Figure 5: Simulation of random variables X and Y such that $\eta = 0.51$

Applications

- The data I am using consists of a 16x16 grid of locations across Lancashire, with rainfall levels recorded every hour from December to February for 25 years.
- I looked into two areas:
  - How distance between points affects the asymptotic dependence between rainfall
  - How asymptotic dependence changes as the aggregation level increases
- Here it can be seen that as distance increases, the values of chi decrease at the same rate for both levels.

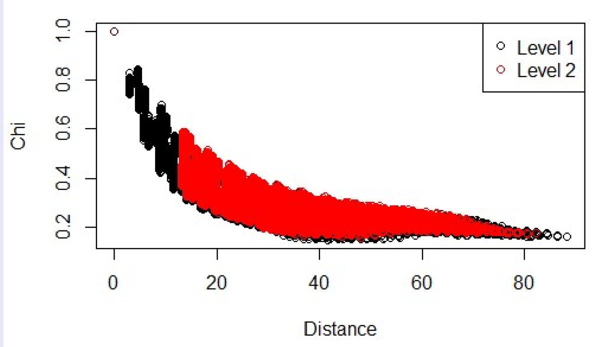

Figure 6: Values of χ for aggregation levels 1 and 2 with respect to distance

References

- Stuart Coles, Janet Heffernan, and Jonathan Tawn (2000) *Dependence Measure for Extreme Value Analyses*
- Stuart Coles (2001) *An Introduction to Statistical Modelling of Extreme Values*

d.bahia@lancaster.ac.uk