

# High-resolution estimation of future extreme rainfall over the UK

Ben Youngman  
[b.youngman@exeter.ac.uk](mailto:b.youngman@exeter.ac.uk)  
<https://github.com/byoungman>

University of Exeter

Statistics seminar  
University of Bath  
14 November 2025

# Motivation

Boscastle floods, August 2004



Pic by Nick Gregory, 16/08/2004.  
Boscastle Flood

# Motivation

## The Great Storm of '87 (87J)



# Motivation

UK heatwave, July 2022



# Motivation

UK coldwave, aka 'The Beast from the East' February 2018



# Overview

- Extreme value theory: a pictorial introduction
- Extreme UK rainfall
- GMRF models for extremes
- Tailored GMRF models for extremes
- Past and future estimates of extreme UK rainfall
- Summary

# Extreme value theory

## The extremal types theorem

- Let  $X_1, \dots, X_n$  be i.i.d. random variables
- Define  $M_n = \max(X_1, \dots, X_n)$
- If there exist sequences of constants  $a_n > 0$  and  $b_n$  such that, as  $n \rightarrow \infty$ ,

$$\Pr\{(M_n - b_n)/a_n \leq x\} \rightarrow H(x)$$

for some non-degenerate distribution  $H$ , then  $H$  is the generalised extreme value (GEV) distribution, denoted  $GEV(\mu, \psi, \xi)$ , where

$$H(x) = \exp \left[ - \left( 1 + \xi \frac{x - \mu}{\psi} \right)_+^{-1/\xi} \right]$$

for  $\{x : (1 + \xi(x - \mu)/\psi) > 0\}$  and where  $\psi > 0$ ,  $(x)_+ = \max(0, x)$  and the case  $\xi = 0$  gives  $H(x) = \exp[-\exp\{-(x - \mu)/\psi\}]$

i.i.d. := "independent and identically distributed". Ref's: Fisher and Tippett (1928); Gnedenko (1943); Jenkinson (1955)

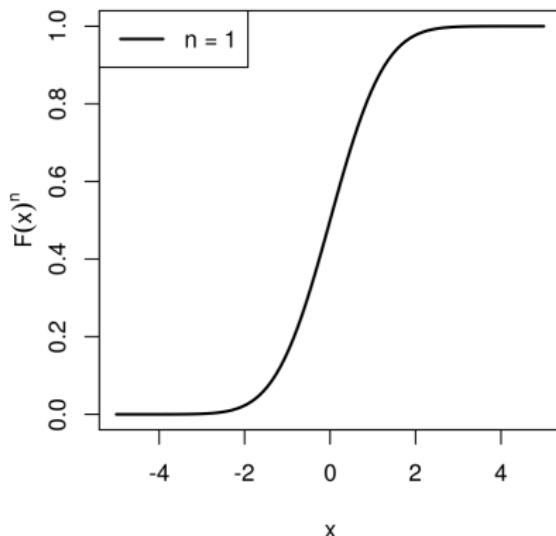
# Extreme value theory

## Normal example

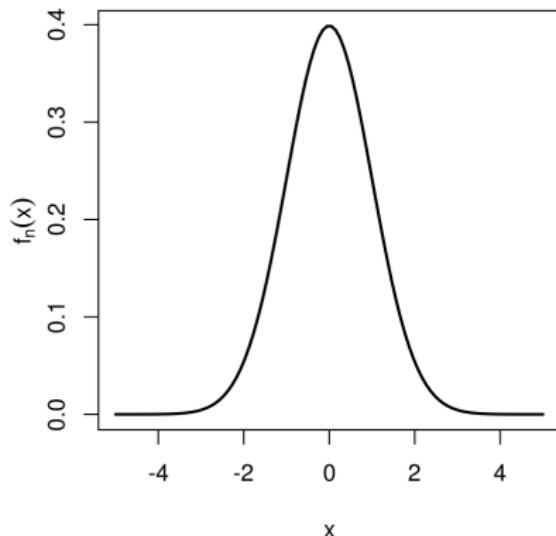
- Suppose  $X_1, \dots, X_n \stackrel{i.i.d.}{\sim} N(0, 1)$
- $\Pr(M_n \leq x) = F(x)^n$

Maxima of normals

Cumulative distribution function



Probability density function

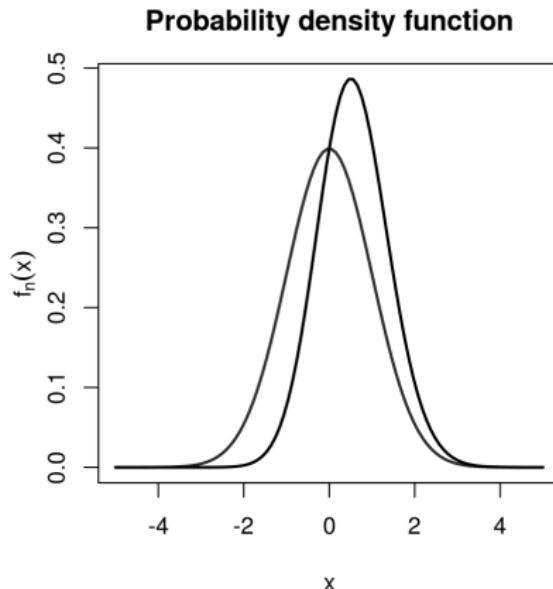
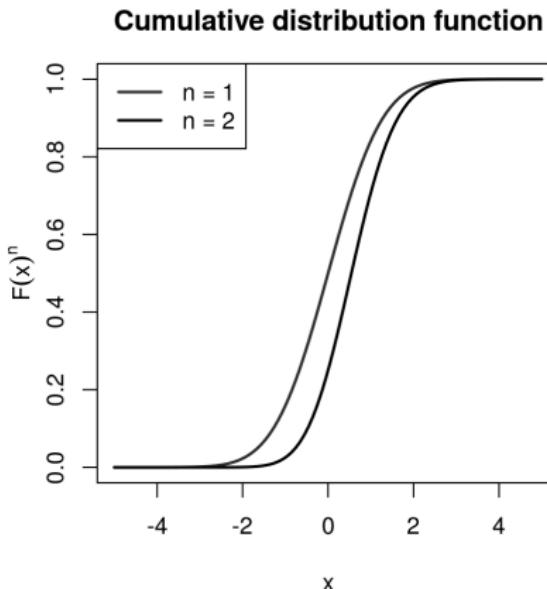


# Extreme value theory

## Normal example

- Suppose  $X_1, \dots, X_n \stackrel{i.i.d.}{\sim} N(0, 1)$
- $\Pr(M_n \leq x) = F(x)^n$

Maxima of normals

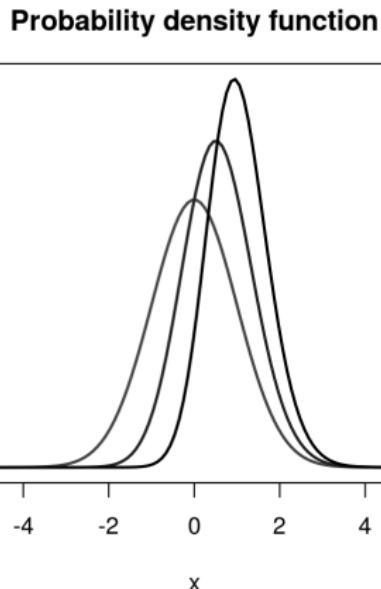
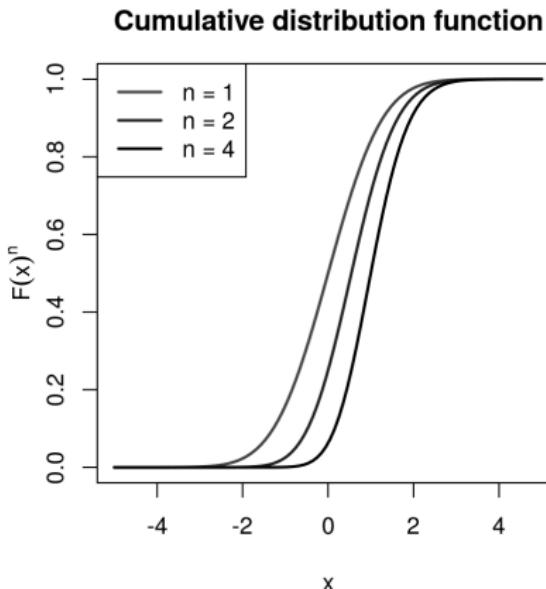


# Extreme value theory

## Normal example

- Suppose  $X_1, \dots, X_n \stackrel{i.i.d.}{\sim} N(0, 1)$
- $\Pr(M_n \leq x) = F(x)^n$

Maxima of normals

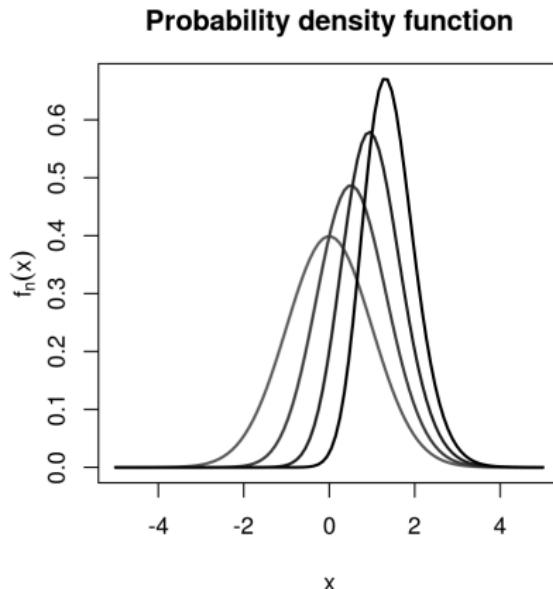
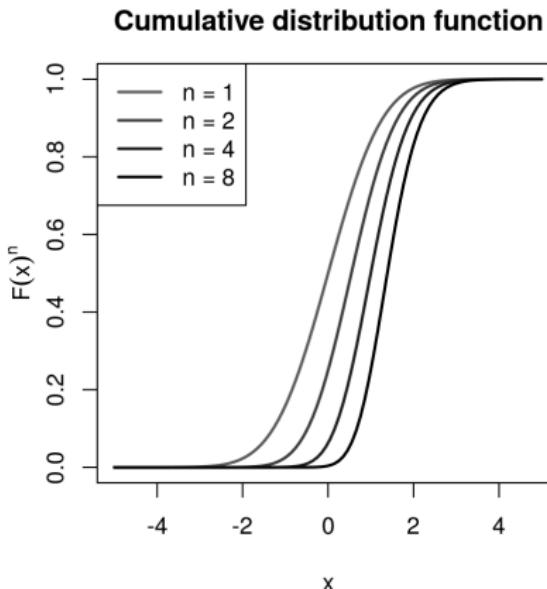


# Extreme value theory

## Normal example

- Suppose  $X_1, \dots, X_n \stackrel{i.i.d.}{\sim} N(0, 1)$
- $\Pr(M_n \leq x) = F(x)^n$

Maxima of normals

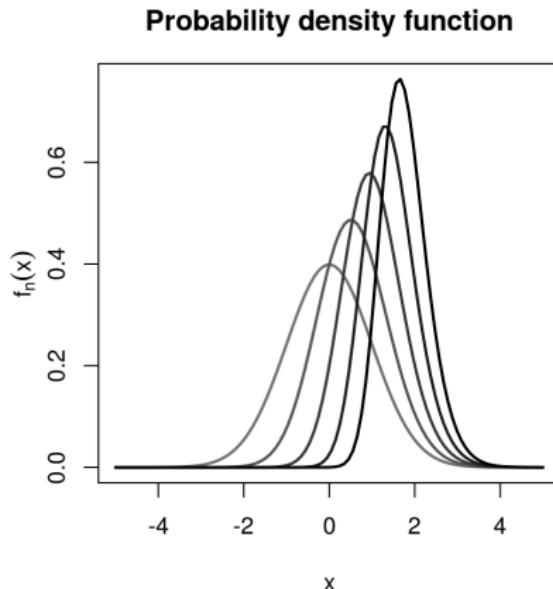
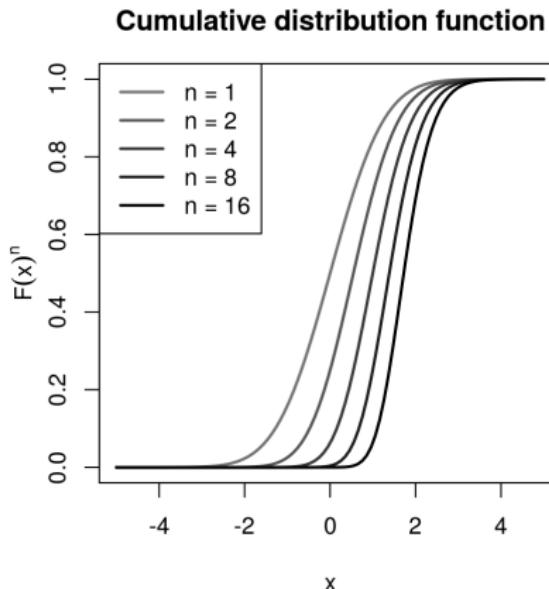


# Extreme value theory

## Normal example

- Suppose  $X_1, \dots, X_n \stackrel{i.i.d.}{\sim} N(0, 1)$
- $\Pr(M_n \leq x) = F(x)^n$

Maxima of normals

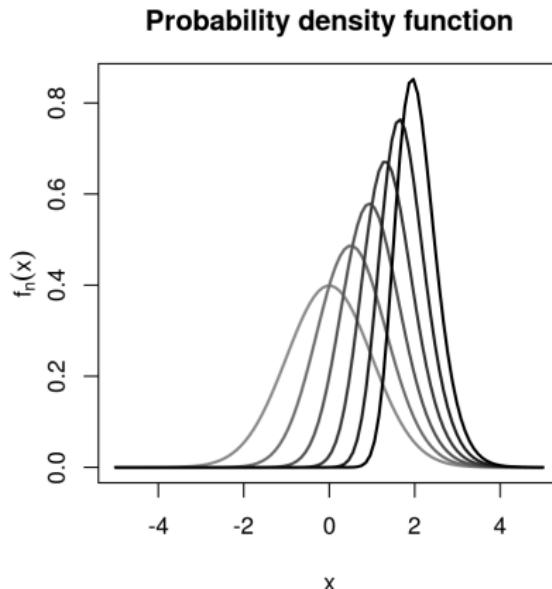
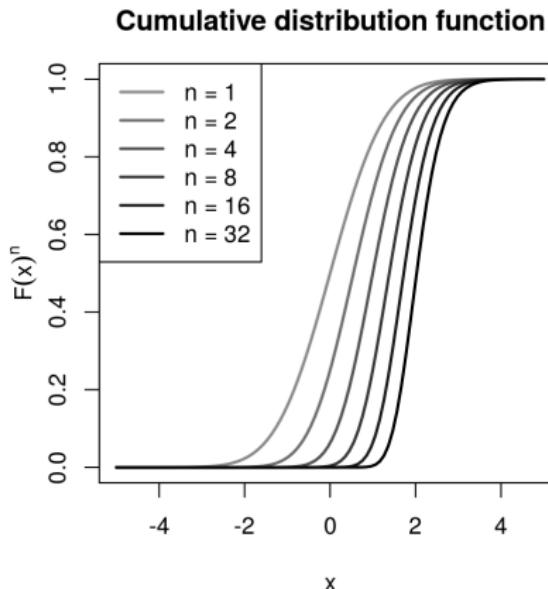


# Extreme value theory

## Normal example

- Suppose  $X_1, \dots, X_n \stackrel{i.i.d.}{\sim} N(0, 1)$
- $\Pr(M_n \leq x) = F(x)^n$

Maxima of normals

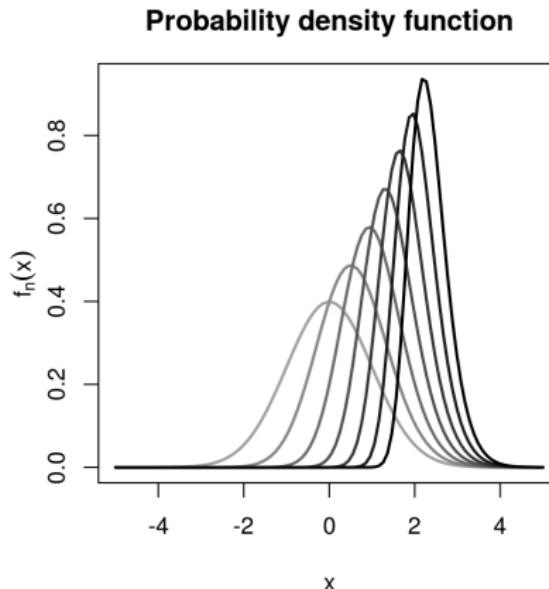
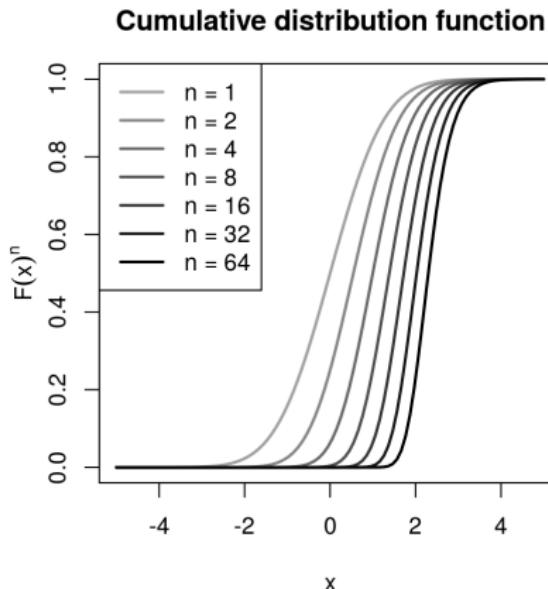


# Extreme value theory

Normal example

- Suppose  $X_1, \dots, X_n \stackrel{i.i.d.}{\sim} N(0, 1)$
- $\Pr(M_n \leq x) = F(x)^n$

Maxima of normals

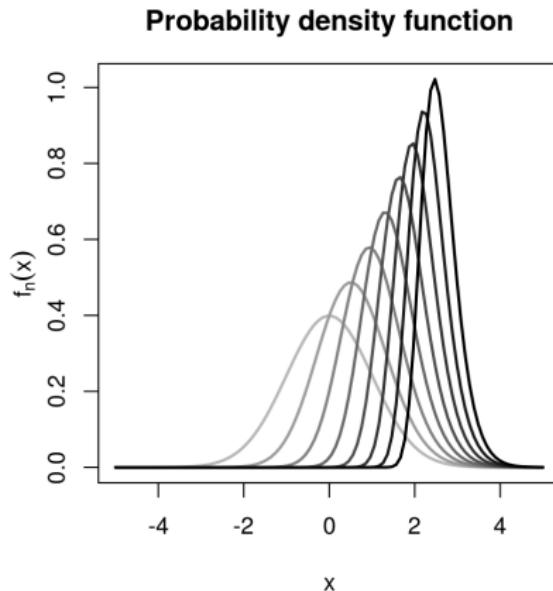
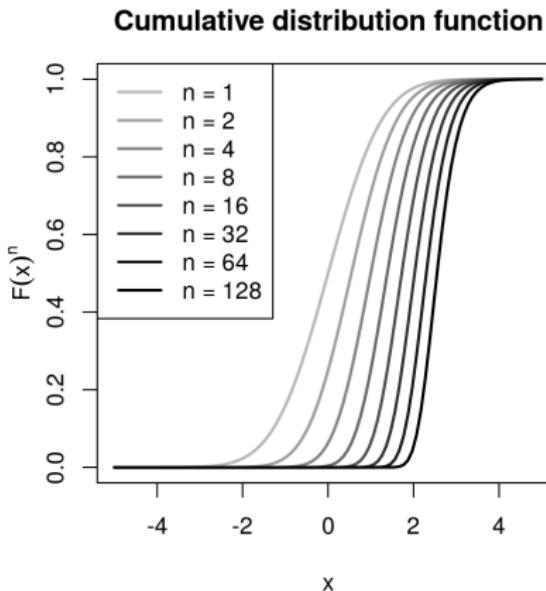


# Extreme value theory

## Normal example

- Suppose  $X_1, \dots, X_n \stackrel{i.i.d.}{\sim} N(0, 1)$
- $\Pr(M_n \leq x) = F(x)^n$

Maxima of normals

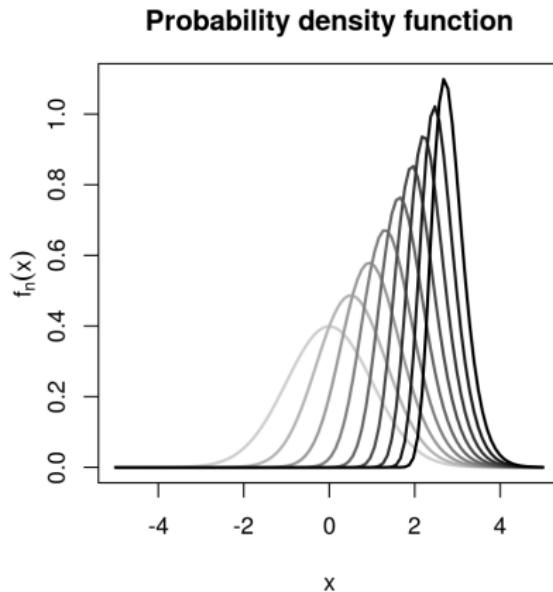
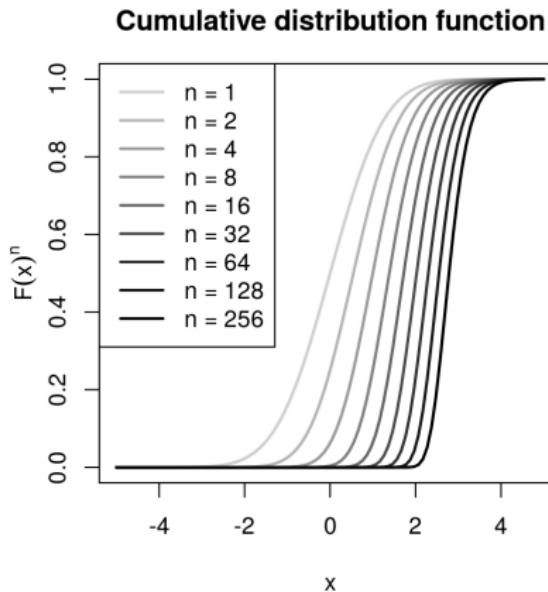


# Extreme value theory

Normal example

- Suppose  $X_1, \dots, X_n \stackrel{i.i.d.}{\sim} N(0, 1)$
- $\Pr(M_n \leq x) = F(x)^n$

Maxima of normals

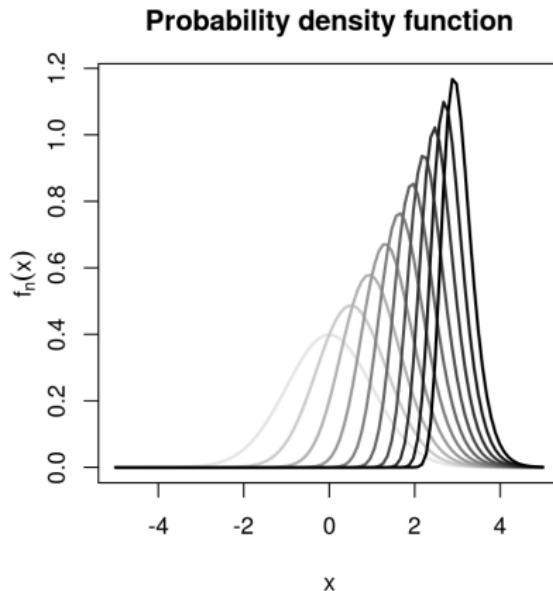
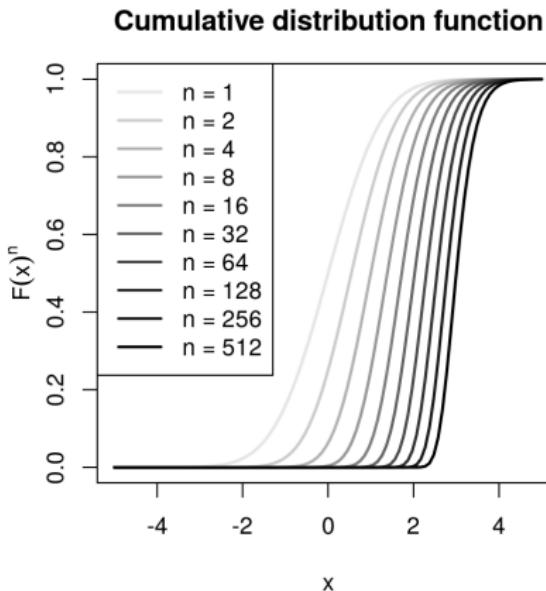


# Extreme value theory

## Normal example

- Suppose  $X_1, \dots, X_n \stackrel{i.i.d.}{\sim} N(0, 1)$
- $\Pr(M_n \leq x) = F(x)^n$

Maxima of normals

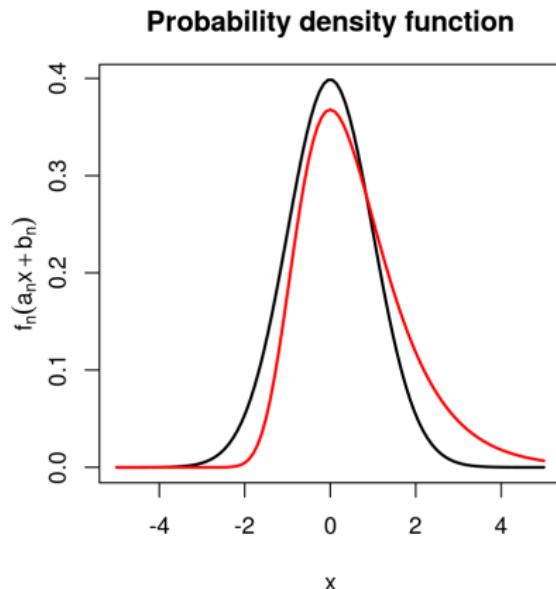
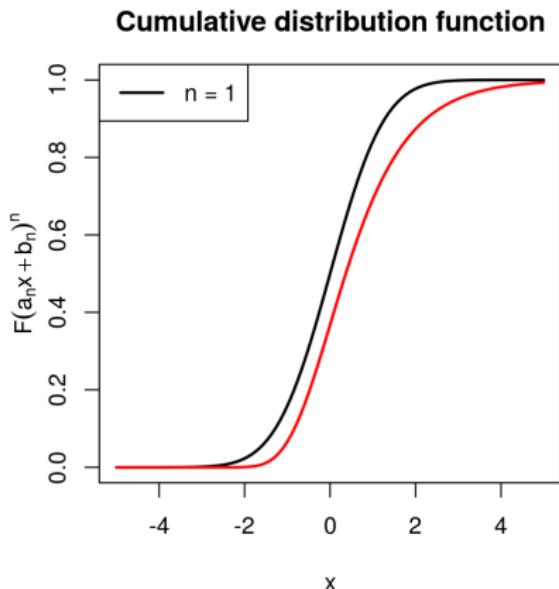


# Extreme value theory

## Normal example

- $\Pr\{(M_n - b_n)/a_n \leq x\} = F(a_n x + b_n)^n$
- $a_n = (2 \log n)^{-\frac{1}{2}}$  and  $b_n = a_n^{-1} - \frac{1}{2}a_n(\log \log n + \log 4\pi)$

Maxima of re-scaled normals

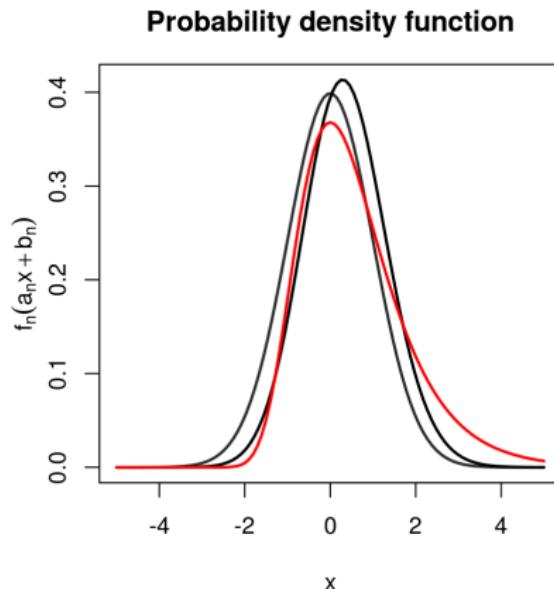
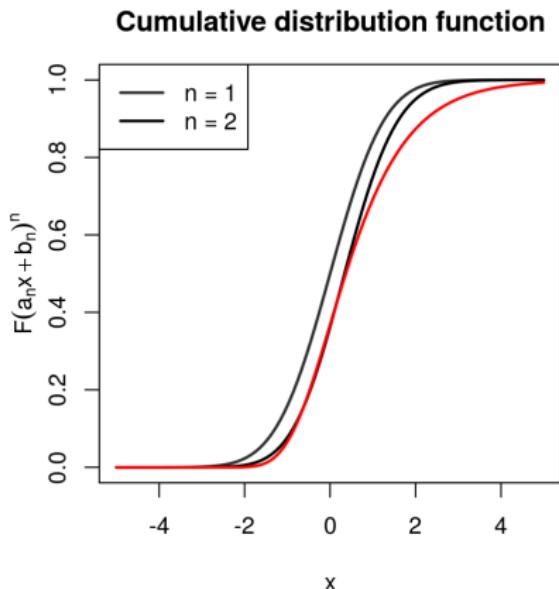


# Extreme value theory

## Normal example

- $\Pr\{(M_n - b_n)/a_n \leq x\} = F(a_n x + b_n)^n$
- $a_n = (2 \log n)^{-\frac{1}{2}}$  and  $b_n = a_n^{-1} - \frac{1}{2}a_n(\log \log n + \log 4\pi)$

Maxima of re-scaled normals

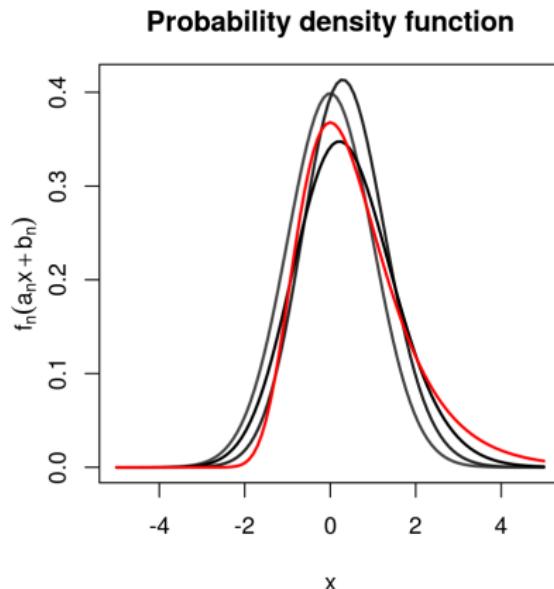
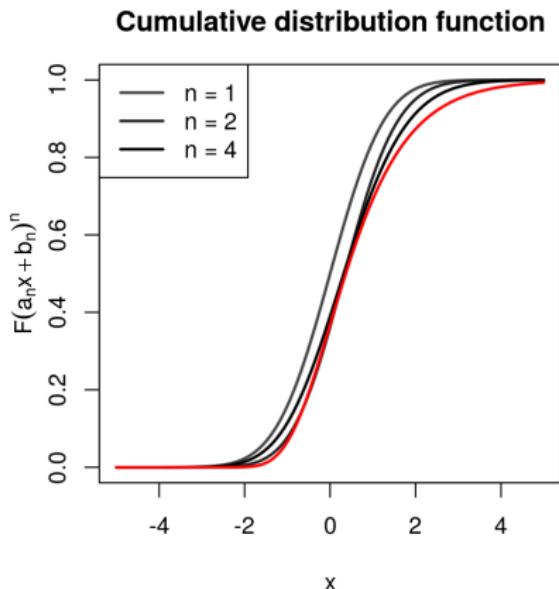


# Extreme value theory

## Normal example

- $\Pr\{(M_n - b_n)/a_n \leq x\} = F(a_n x + b_n)^n$
- $a_n = (2 \log n)^{-\frac{1}{2}}$  and  $b_n = a_n^{-1} - \frac{1}{2}a_n(\log \log n + \log 4\pi)$

Maxima of re-scaled normals

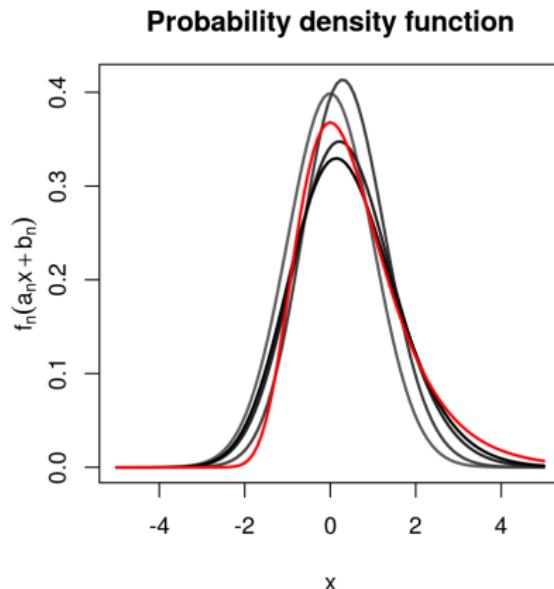
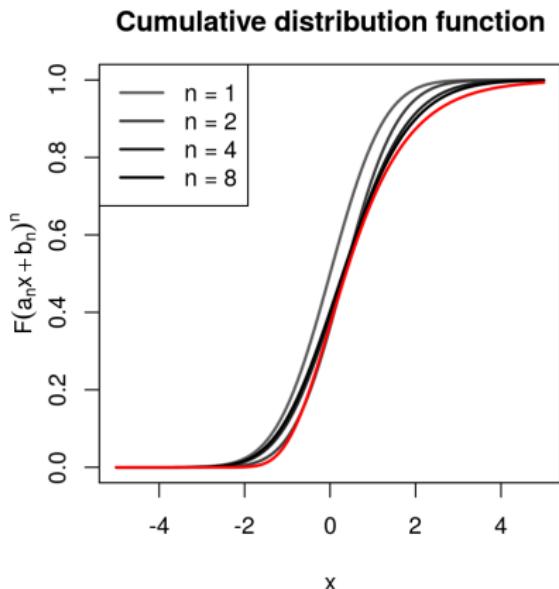


# Extreme value theory

## Normal example

- $\Pr\{(M_n - b_n)/a_n \leq x\} = F(a_n x + b_n)^n$
- $a_n = (2 \log n)^{-\frac{1}{2}}$  and  $b_n = a_n^{-1} - \frac{1}{2}a_n(\log \log n + \log 4\pi)$

Maxima of re-scaled normals

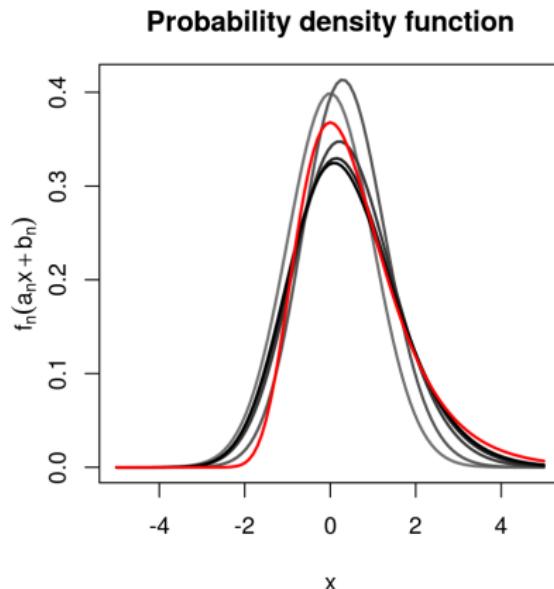
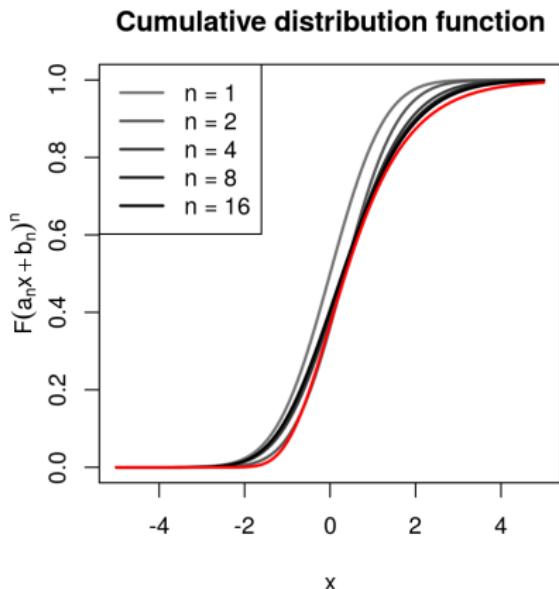


# Extreme value theory

## Normal example

- $\Pr\{(M_n - b_n)/a_n \leq x\} = F(a_n x + b_n)^n$
- $a_n = (2 \log n)^{-\frac{1}{2}}$  and  $b_n = a_n^{-1} - \frac{1}{2}a_n(\log \log n + \log 4\pi)$

Maxima of re-scaled normals

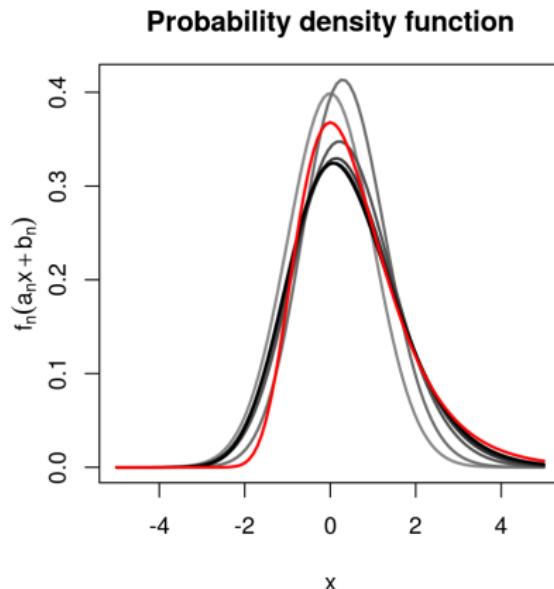
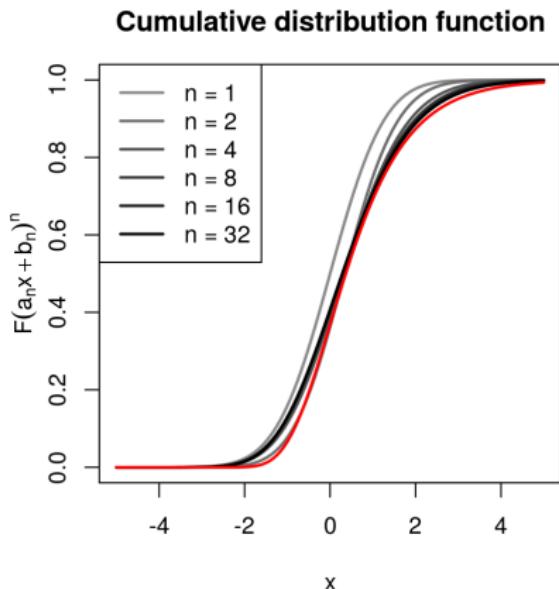


# Extreme value theory

## Normal example

- $\Pr\{(M_n - b_n)/a_n \leq x\} = F(a_n x + b_n)^n$
- $a_n = (2 \log n)^{-\frac{1}{2}}$  and  $b_n = a_n^{-1} - \frac{1}{2}a_n(\log \log n + \log 4\pi)$

Maxima of re-scaled normals

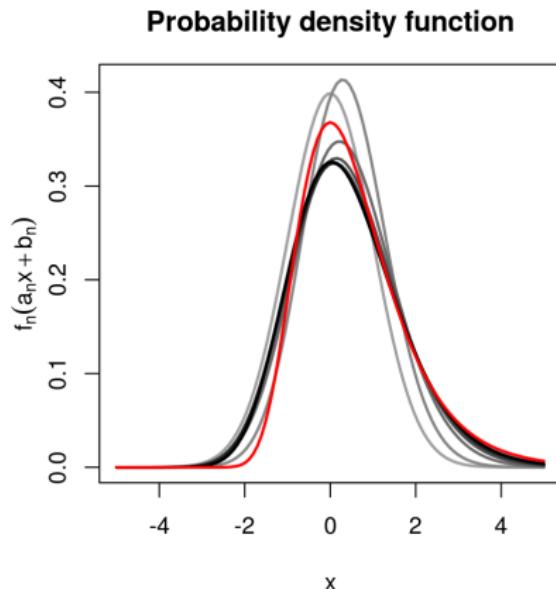
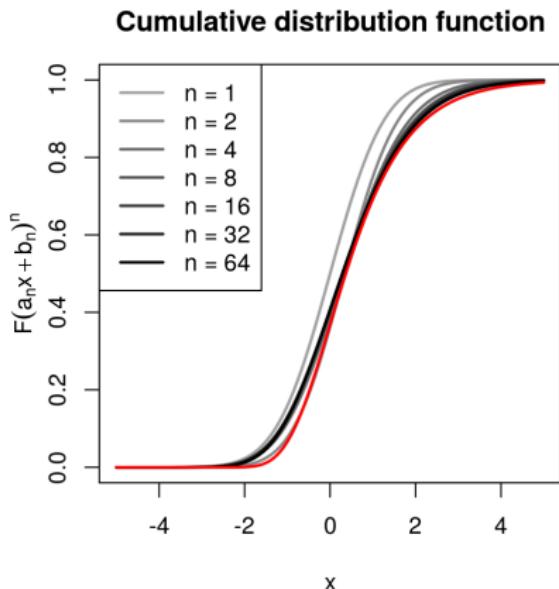


# Extreme value theory

## Normal example

- $\Pr\{(M_n - b_n)/a_n \leq x\} = F(a_n x + b_n)^n$
- $a_n = (2 \log n)^{-\frac{1}{2}}$  and  $b_n = a_n^{-1} - \frac{1}{2}a_n(\log \log n + \log 4\pi)$

Maxima of re-scaled normals

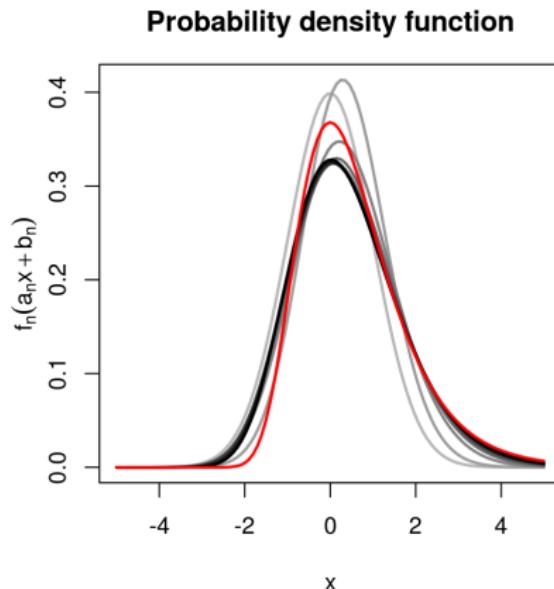
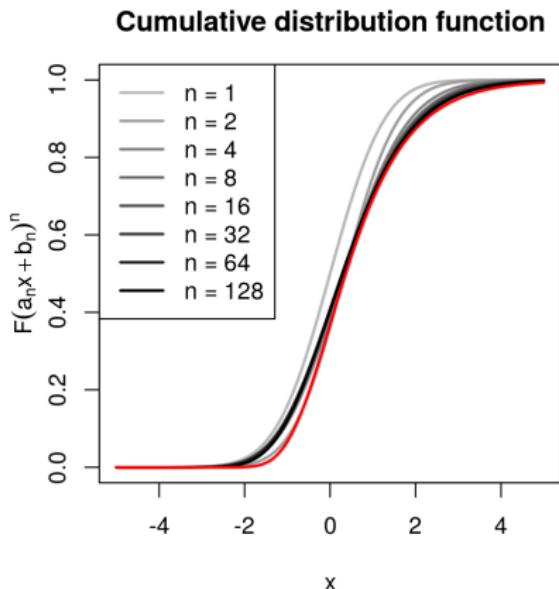


# Extreme value theory

## Normal example

- $\Pr\{(M_n - b_n)/a_n \leq x\} = F(a_n x + b_n)^n$
- $a_n = (2 \log n)^{-\frac{1}{2}}$  and  $b_n = a_n^{-1} - \frac{1}{2}a_n(\log \log n + \log 4\pi)$

Maxima of re-scaled normals

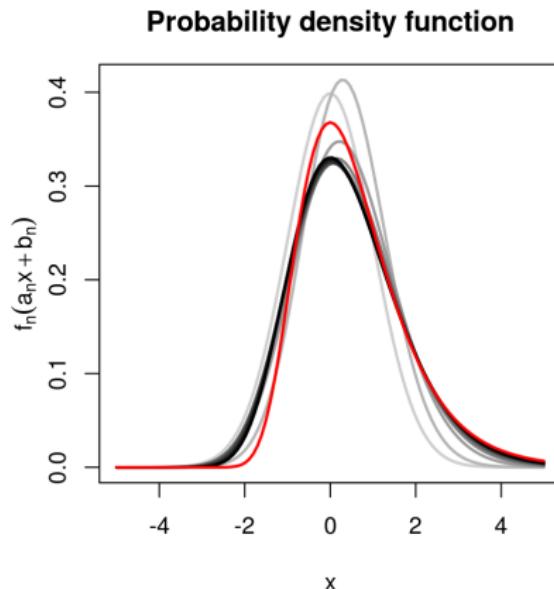
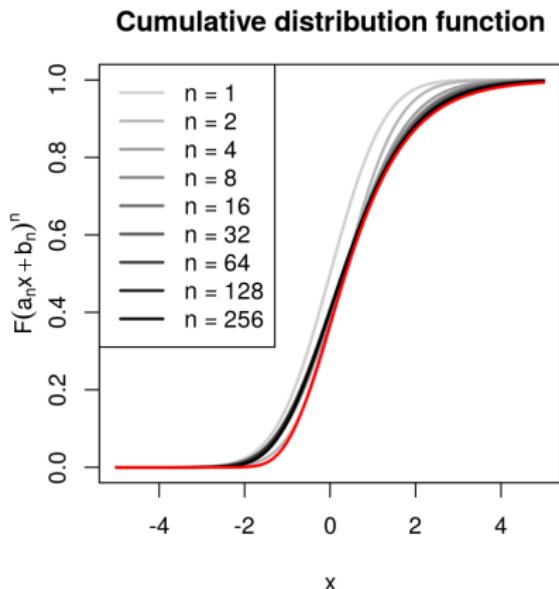


# Extreme value theory

## Normal example

- $\Pr\{(M_n - b_n)/a_n \leq x\} = F(a_n x + b_n)^n$
- $a_n = (2 \log n)^{-\frac{1}{2}}$  and  $b_n = a_n^{-1} - \frac{1}{2}a_n(\log \log n + \log 4\pi)$

Maxima of re-scaled normals

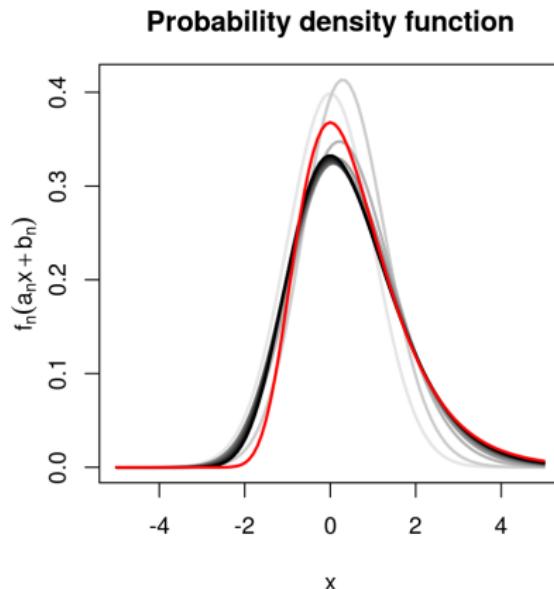
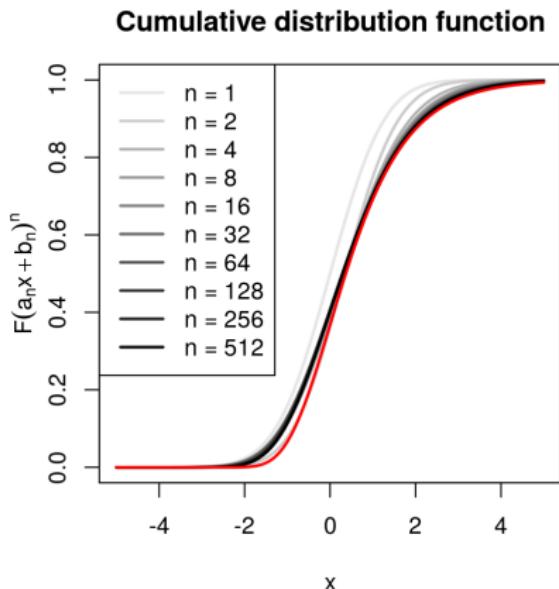


# Extreme value theory

## Normal example

- $\Pr\{(M_n - b_n)/a_n \leq x\} = F(a_n x + b_n)^n$
- $a_n = (2 \log n)^{-\frac{1}{2}}$  and  $b_n = a_n^{-1} - \frac{1}{2}a_n(\log \log n + \log 4\pi)$

Maxima of re-scaled normals



# Statistical models for maxima

## block maxima/minima

- Let  $X_1, X_2, \dots, X_{nm}$  be an i.i.d. sequence
- Partition data into  $n$  blocks of length  $m$ , i.e.,

$$\{X_{jm+1}, \dots, X_{jm+m}\}_{j=0, \dots, n-1}$$

- Obtain *block maxima*

$$Y_j = \max\{X_{jm+1}, \dots, X_{jm+m}\}$$

- For  $j = 1, \dots, n$  assume

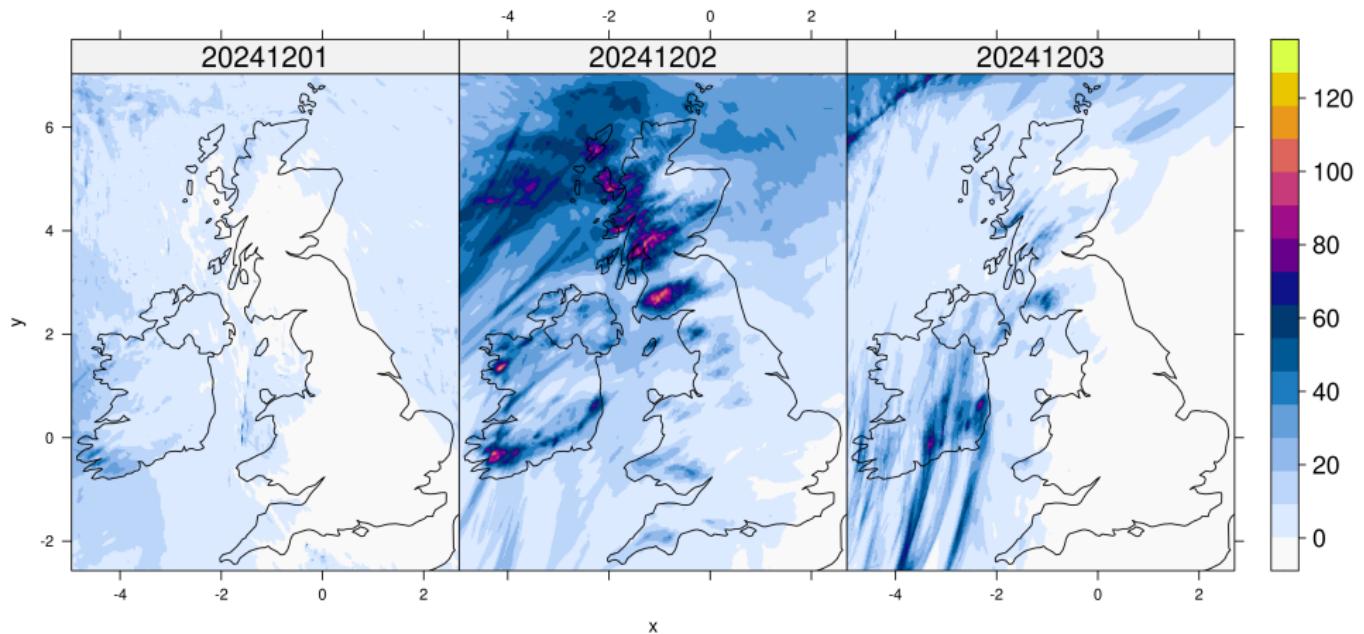
$$Y_j \sim GEV(\mu, \psi, \xi)$$

- Model *minima* by noting that

$$\min(X_1, X_2, \dots) = -\max(-X_1, -X_2, \dots)$$

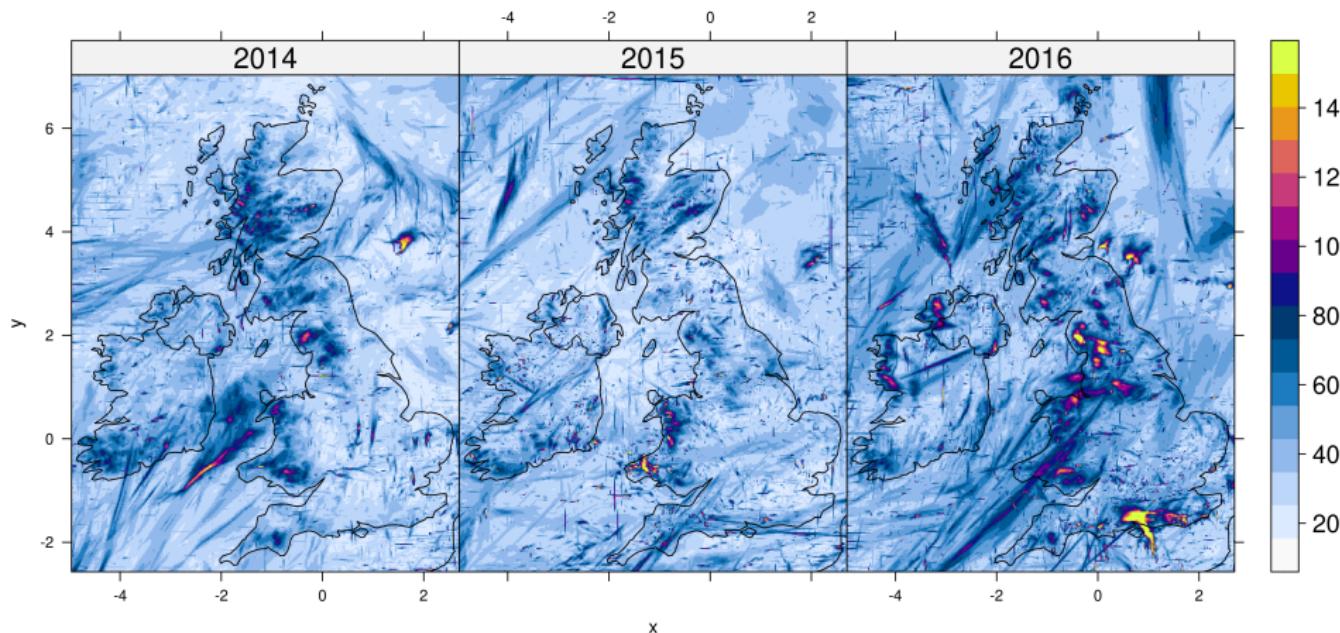
# High resolution UK rainfall projections

- The UKCP18 project gives daily rainfall projections at 2.2km resolution
  - it's the best available, at the moment



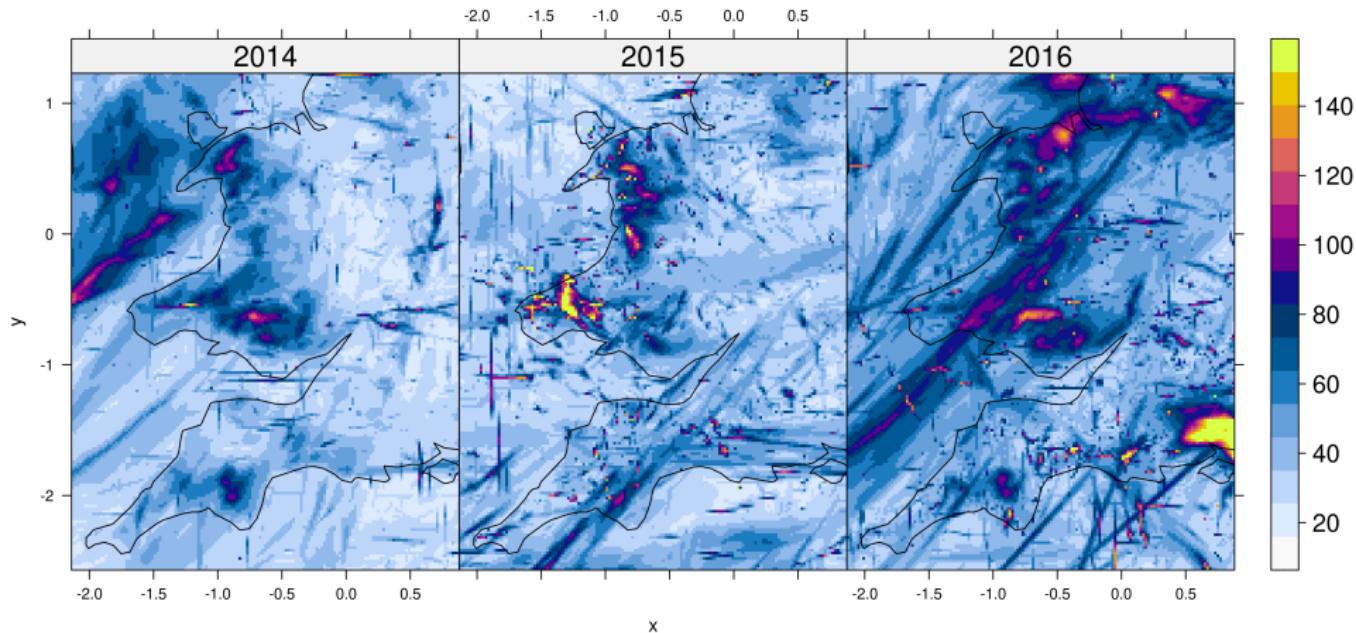
# High resolution UK rainfall projections

- We want to work with annual maxima



# High resolution UK rainfall projections

- But numerical errors risk dominating our analysis
- We attempt to filter out obvious horizontal and vertical lines



## Modelling annual maxima

- Let  $Y_{ijt}$  denote the annual maximum in year  $t$  for grid row  $i$  and column  $j$
- We have a  $318 \times 516$  grid ( $\sim 200,000$  grid cells) and 20 years of data
- We assume

$$Y_{ijt} \sim GEV(\mu_{ij}, \psi_{ij}, \xi_{ij})$$

- We could consider grid cells as spatial locations, e.g.  $s_{ij}$ , and model with continuous processes, such as  $\mu_{ij} = \mu(s_{ij})$ , via smooth spatial processes (e.g. Gaussian processes or thin-plate splines)
  - but even with covariate information, we're likely to impose unwanted assumptions, especially over large domain
- A local smoother, such as a Gaussian Markov random field (GMRF), might work better
  - examples include Cooley and Sain (2010) and Auld et al. (2023), amongst others

## Gaussian Markov random field (GMRF)

- Consider a random vector  $\mathbf{Z} = (Z_1, \dots, Z_n)$
- With a GMRF model we assume that

$$\mathbf{Z} \sim N_n(\mathbf{m}, \mathbf{Q}^{-1})$$

i.e. is  $n$ -dimensional multivariate Gaussian with precision matrix  $\mathbf{Q}$

- The conditional expectation is

$$E(Z_i | z_j, j \neq i) = m_i - \frac{\sum_{j \neq i} Q_{ij}(z_j - m_j)}{Q_{ii}}$$

- So if  $Q_{ij} = 0$ , then  $Z_i$  and  $Z_j$  are conditionally independent
- Or  $Z_i$  and  $Z_j$  are neighbours if  $Q_{ij} \neq 0$

## GMRFs on grids

- Consider the collections of GEV parameters  $\mu$ ,  $\psi$  and  $\xi$
- We use GMRFs to specify conditional expectations amongst GEV parameters
- Now consider  $Z_{ij}$  on the grid  $(i, j)$  for  $i = 1, \dots, n_x$  and  $j = 1, \dots, n_y$ .
- A simple GMRF assumes  $Z_{ij}$  is the average of its four nearest neighbours,  $Z_{i-1,j}, Z_{i+1,j}, Z_{i,j-1}, Z_{i,j+1}$ , so that

$$E(Z_{ij} \mid z_{i-1,j}, z_{i+1,j}, z_{i,j-1}, z_{i,j+1}) = \frac{1}{4}(z_{i-1,j} + z_{i+1,j} + z_{i,j-1} + z_{i,j+1})$$

- This gives an intrinsic GMRF for a regular grid

## Intrinsic GMRF precision matrices

- The intrinsic GMRF precision matrix can be written as

$$\mathbf{Q}_n^{(q)} = \mathbf{I}_{n_y} \otimes \mathbf{L}_{n_x} + \mathbf{L}_{n_y} \otimes \mathbf{I}_{n_x}$$

where

$$\mathbf{L}_m = \mathbf{D}_m^\top \mathbf{D}_m$$

with  $m \times m$  matrix  $\mathbf{D}_m$  of the form

$$\mathbf{D}_m = \begin{bmatrix} -1 & 1 & & & \\ & -1 & 1 & & \\ & & \ddots & \ddots & \\ & & & -1 & 1 \end{bmatrix}$$

- This construction easily extends to higher-order precision matrices, which corresponds to smoothing over more neighbours

## Aside: GMRFs on graphs

- An alternative construction is  $\mathbf{Q} = \kappa(\mathbf{D} - \mathbf{W})$
- $\mathbf{W}$  is an adjacency matrix with

$$w_{ij} = \begin{cases} 1 & \text{if } i \sim j, \text{ i.e. } i \text{ and } j \text{ are neighbours} \\ 0 & \text{otherwise} \end{cases}$$

- $\mathbf{D} = \text{diag}(d_1, \dots, d_n)$  with  $d_i = \sum_{j=1}^n w_{ij}$
- This general formulation allows for more general spaces than grids to be considered
- Such spaces might want to be considered in terms of graphs and conditional independence

## GMRFs for extreme values

- GEV parameters are assumed to have GMRF distributions

$$\begin{aligned}\boldsymbol{\eta}_\mu = \boldsymbol{\mu} &\sim N_n(\boldsymbol{\mu}, \mathbf{Q}_\mu^{-1}), \quad \boldsymbol{\eta}_\psi = h_\psi(\psi) \sim N_n(\psi, \mathbf{Q}_\psi^{-1}), \\ \boldsymbol{\eta}_\xi = h_\xi(\xi) &\sim N_n(\xi, \mathbf{Q}_\xi^{-1})\end{aligned}$$

where link function  $h_\psi$  ensures  $\psi > 0$  and  $h_\xi$  ensures  $\xi$  in  $[-1, 0.5]$

- Given an appropriate precision matrix  $\mathbf{Q} = \text{diag}(\mathbf{Q}_\mu, \mathbf{Q}_\psi, \mathbf{Q}_\xi)$ , we obtain estimates of  $\boldsymbol{\mu}$ ,  $\boldsymbol{\psi}$  and  $\boldsymbol{\xi}$  by maximising

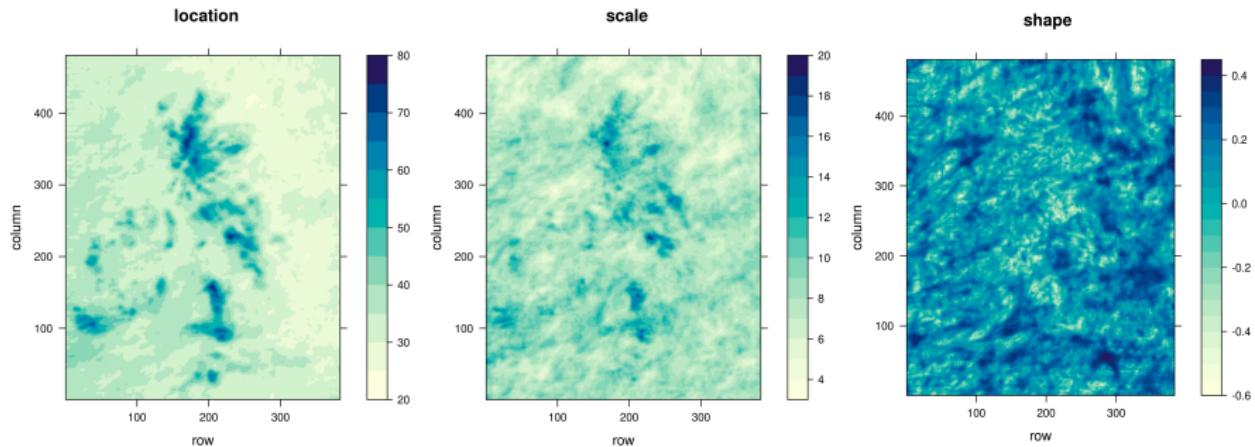
$$\log f(\mathbf{y} \mid \boldsymbol{\mu}, \boldsymbol{\psi}, \boldsymbol{\xi}) - \frac{1}{2} \boldsymbol{\eta}^\top \mathbf{Q} \boldsymbol{\eta}$$

where  $\boldsymbol{\eta} = (\boldsymbol{\eta}_\mu^\top, \boldsymbol{\eta}_\psi^\top, \boldsymbol{\eta}_\xi^\top)^\top$

- However, to estimate any parameters governing  $\mathbf{Q}$ , this needs to be done iteratively using a Laplace approximation; see Wood (2010)

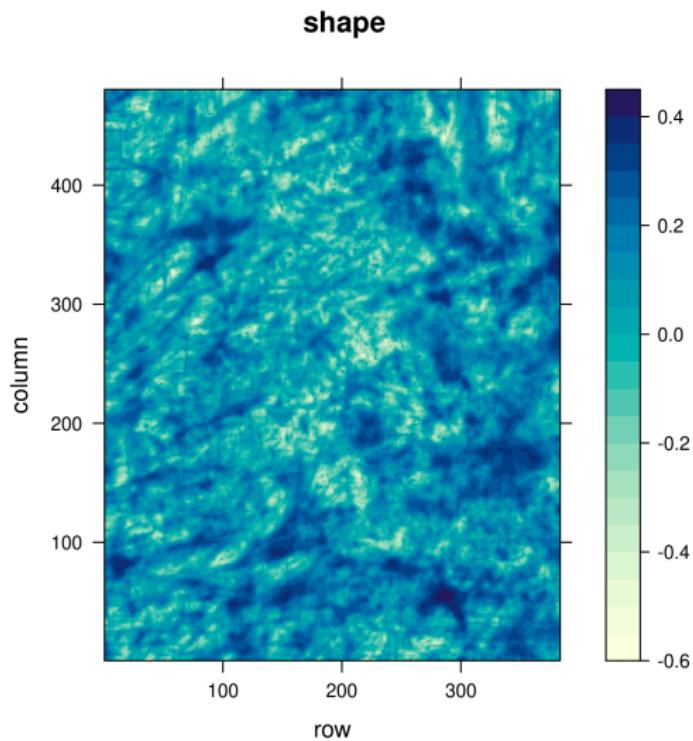
# Model fitting

- GEV parameter estimates are easily mapped



## Model fitting

- But our model cannot smooth out numerical errors

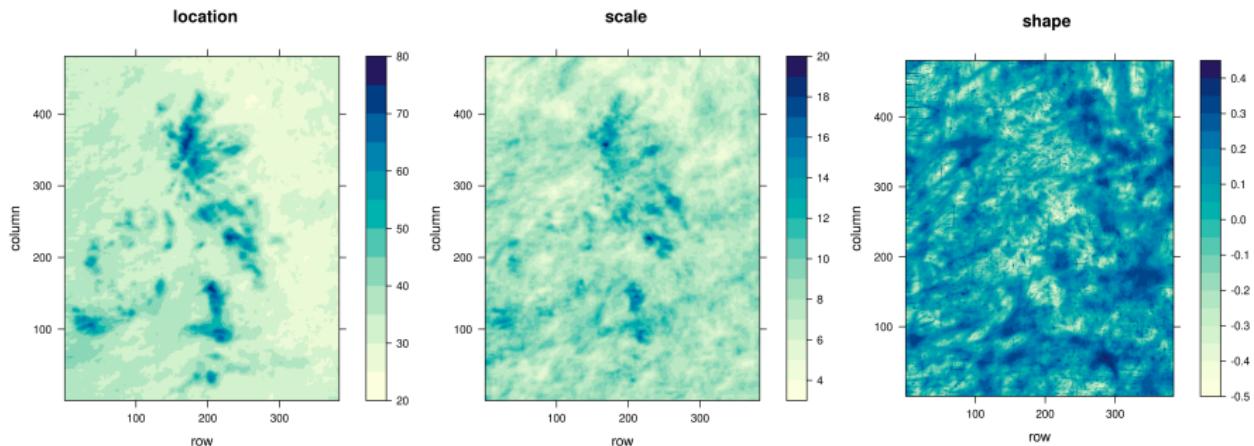


## The Besag-York-Mollie (2) model

- The BYM2 model partitions  $\mathbf{Z}$  as  $\mathbf{Z} = \mathbf{U} + \mathbf{V}$ , where  $\mathbf{U} \sim N_n(u, \mathbf{Q}_u^{-1})$  is a spatial component and  $\mathbf{V} \sim N_n(0, \sigma^2 \mathbf{I})$  are iid random effects
- The extra flexibility of the BYM2 model is at the cost of another  $n + 1$  parameters to estimate
- But its signal + noise representation is particularly appealing when estimating differences in extremes, because  $\mathbf{V}$  can be considered noise, and our analysis can be based on changes in signal alone
- Because  $U_i \sim N(0, \sigma^2)$ , this model gives an intuitive way of specifying random effects to capture erroneous data
- We assume that  $U_i$  have a normal-inverse Gaussian distribution
  - specified to be very close to a  $N(0, \sigma^2)$  distribution, allowing for a larger proportion of values  $U_i \gg 3\sigma$

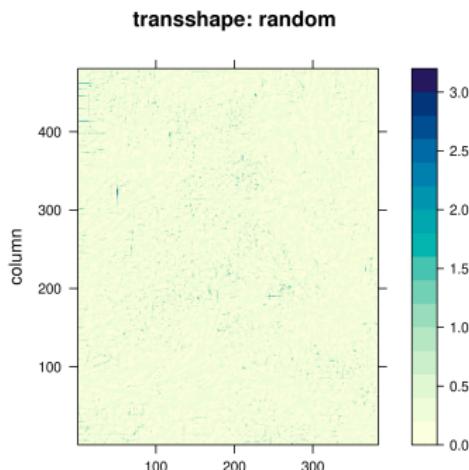
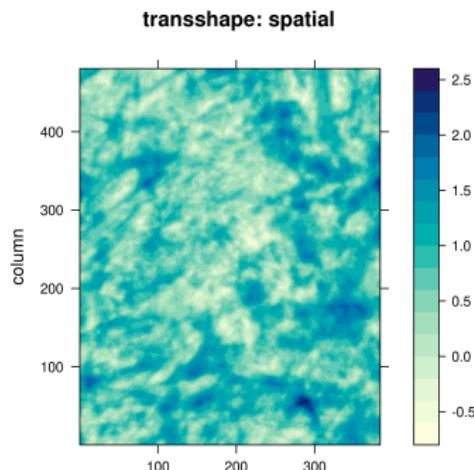
# GMRFs for extreme values with non-Gaussian random effects

- The BYM2 model lets us single out numerical errors



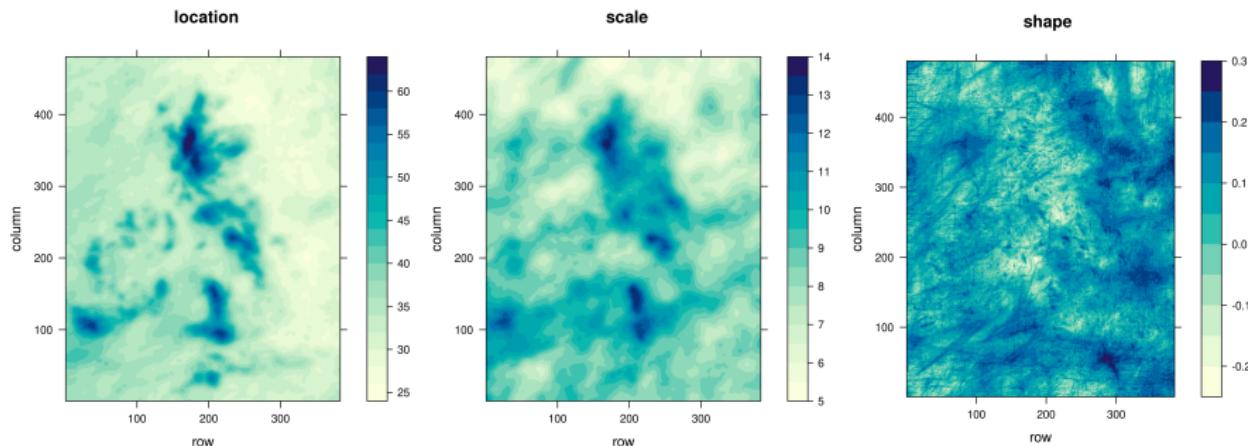
# GMRFs for extreme values with non-Gaussian random effects

- The BYM2 model lets us single out numerical errors



# Miss-specified model adjustments

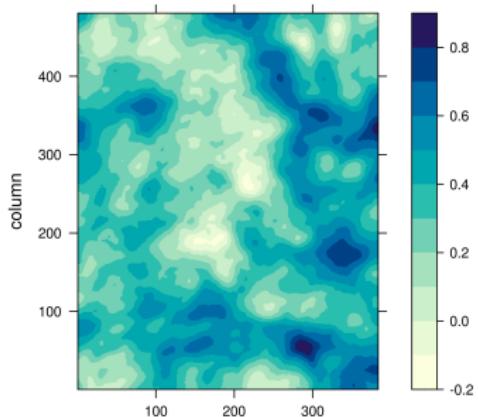
- Our modelling framework assumes that  $Y_i \perp Y_j | (\mu_i, \psi_i, \xi_i, \mu_j, \psi_j, \xi_j)$
- But nearby on the grid  $Y_i$  and  $Y_j$  will often come from the same storm, which results in over-confident estimates
- Using a weighted likelihood and cross-validation we calibrate estimate uncertainty



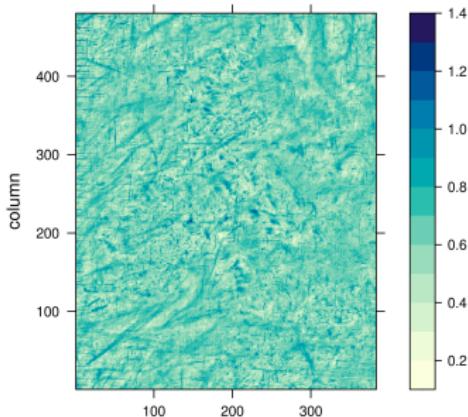
# Mis-specified model adjustments

- Resulting estimates are smoother and better capture spatial variation
- The non-Gaussian random effects still capture local variations and numerical errors

transshape: spatial



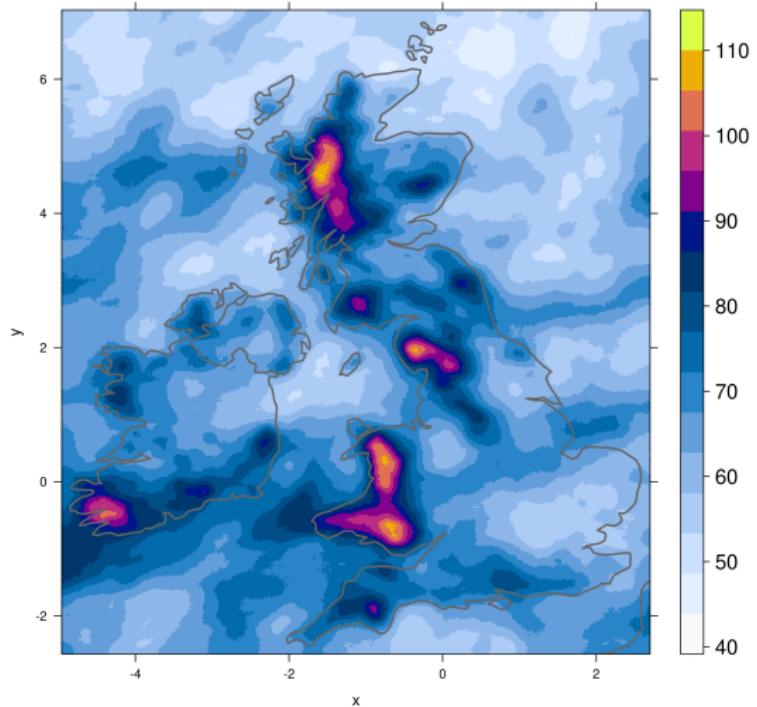
transshape: random



## 100-year return levels estimates

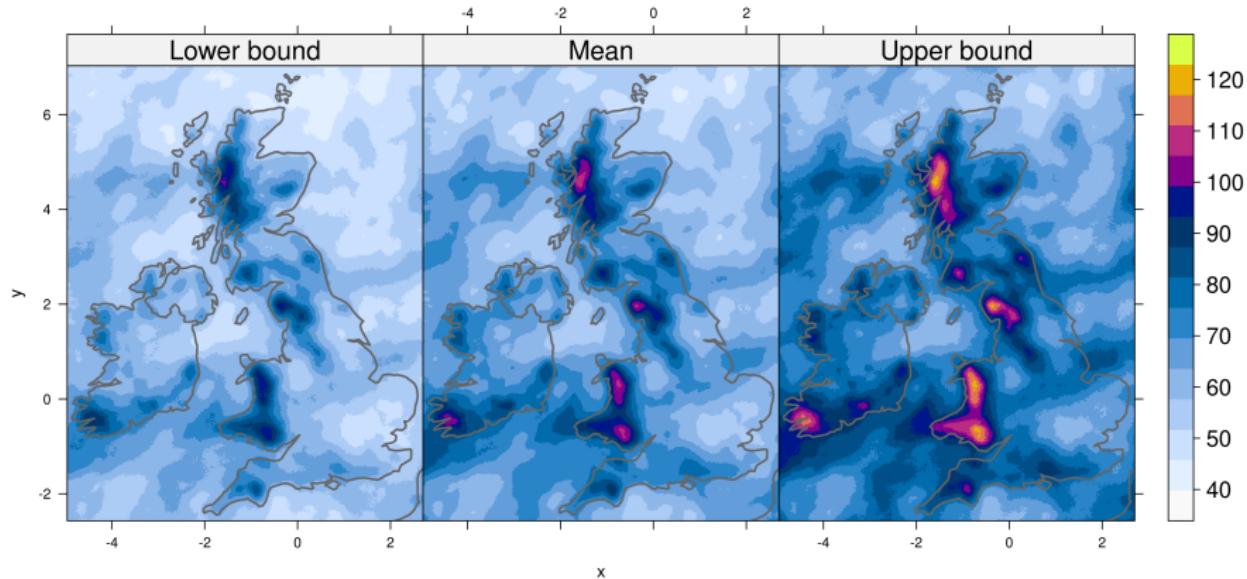
- Because we model annual maxima, the  $p$ -year return level is

$$z_p = \mu - \frac{\psi}{\xi} [1 - \{ -\log(1 - 1/p) \}]^{-\xi}, \quad \xi \neq 0$$



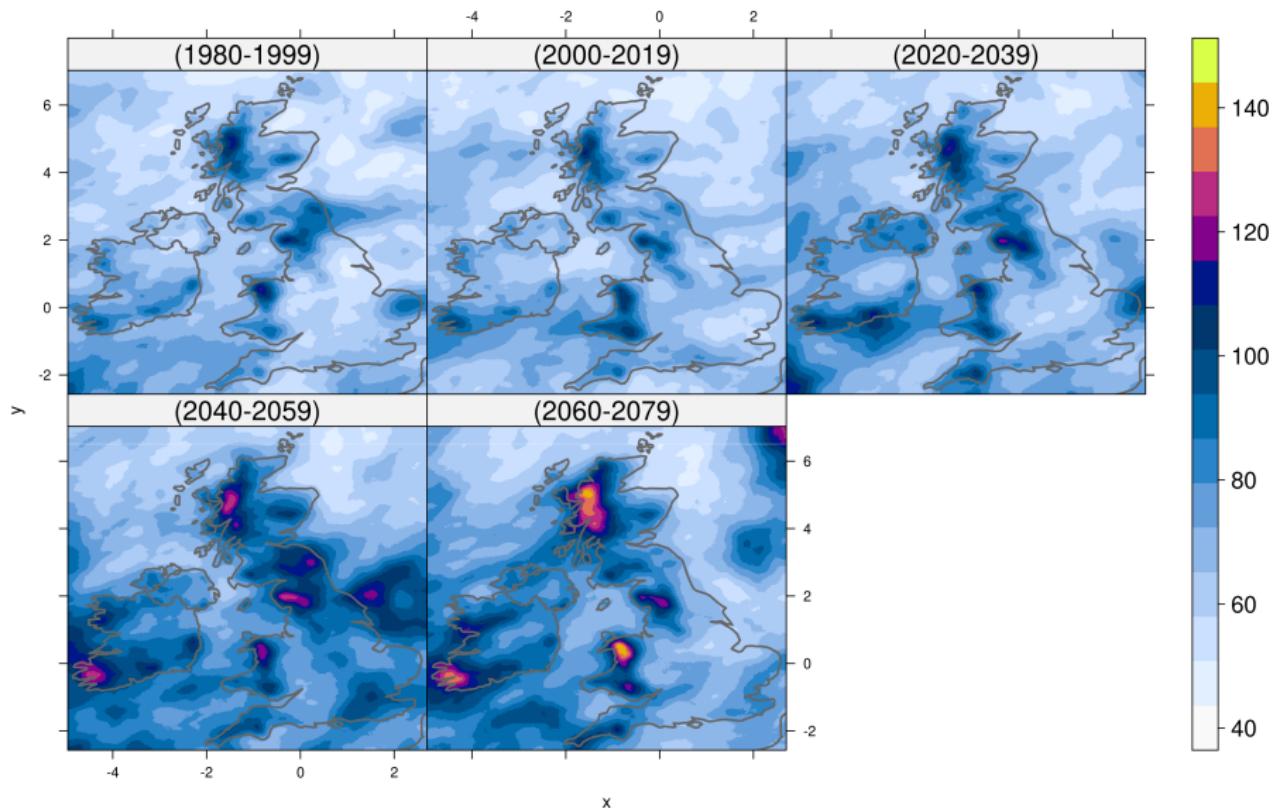
# 100-year return levels estimates with 95% confidence intervals

- We get separate return levels for different 20-year time periods

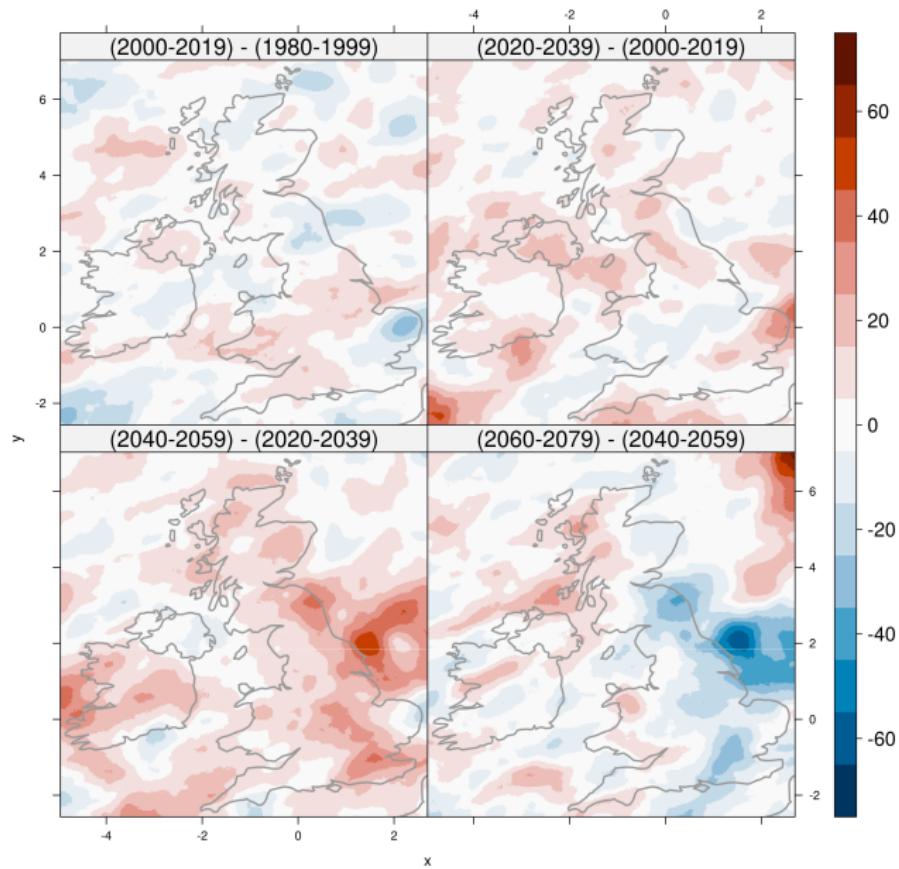


# 100-year return levels estimates over different epochs

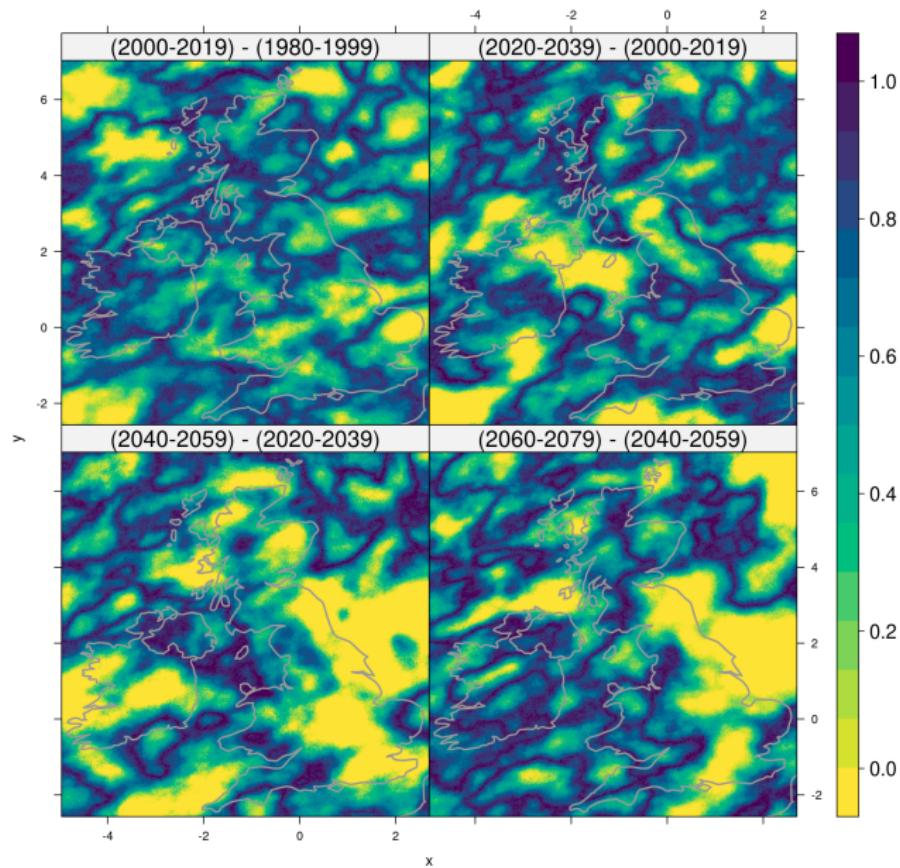
- We get separate return levels for different 20-year time periods



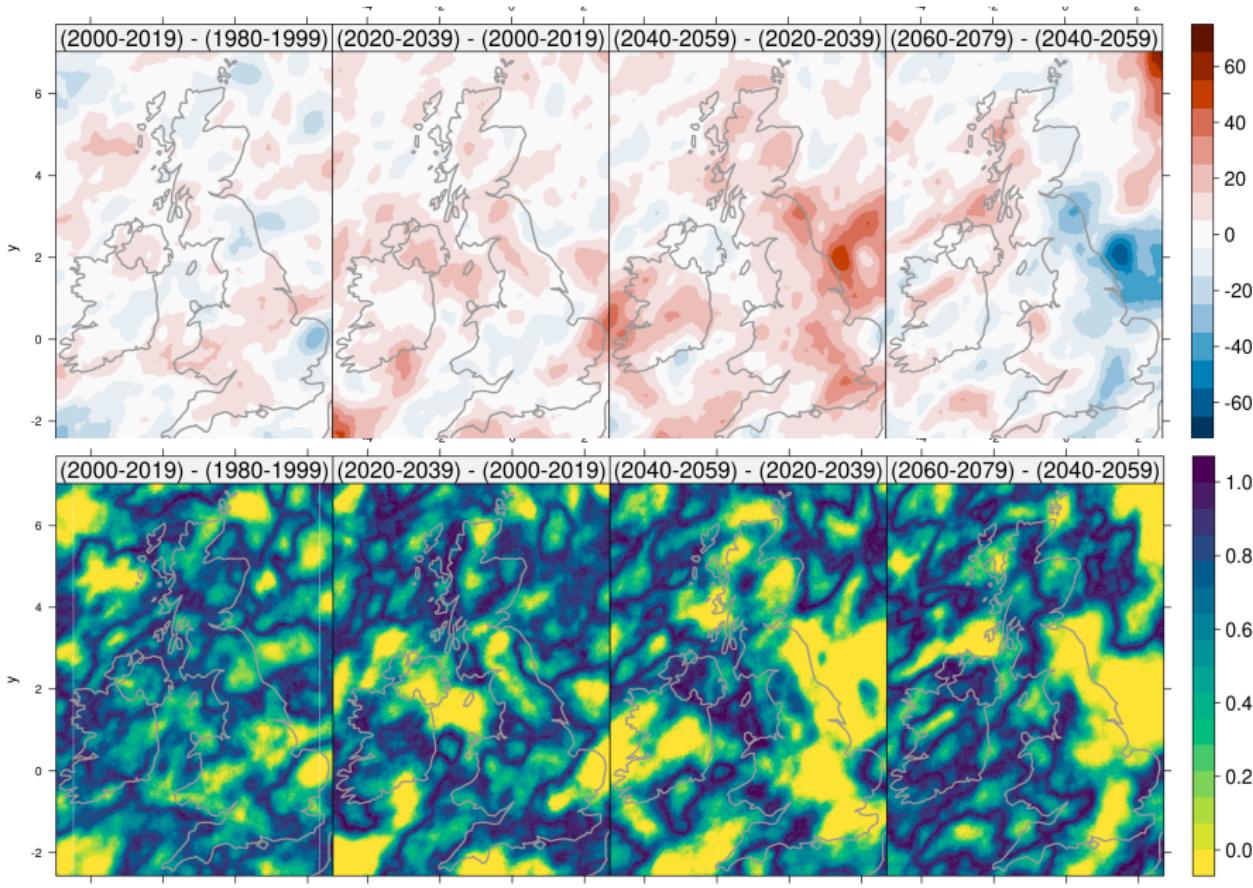
# Changes over epochs



# Confidence interval overlap



# Changes over epochs and confidence interval overlap



## evgmrf for fitting models in R

- We've developed R package evgmrf for (hopefully) straightforward fitting of extreme value models with parameters represented with GMRFs
- It's designed so that users just supply an array and choose their family
  - > mod <- evgmrf(array, family = 'gev')
- > predict(mod, se.fit = FALSE/TRUE) gives parameter estimates, with or without standard errors (se.fit = TRUE), which can be plotted with
- > plot(mod) can then plot estimates
- > predict(mod, prob = p) or plot(mod, prob = p) give/plot extreme value distribution quantiles, which gives easy access to return level estimates
- Available at <https://github.com/byoungman/evgmrf>

## Summary

- GMRF models for extreme value distribution parameters support incredibly high dimensions
- Such models offer a fairly off-the-shelf approach to coherently modelling gridded data
- We demonstrate the use of the models on 2.2km rainfall annual maxima across the UK and Ireland
- New model specifications let us capture numerical errors in the data
- Comparisons across different 20-year periods do not give compelling evidence of changes
- Using annual maxima is rather wasteful of data: investigations suggest the  $r$ -largest order statistics model might be more robust at detecting changes

## References |

- Auld, G., G. C. Hegerl, and I. Papastathopoulos (2023). Changes in the distribution of annual maximum temperates in europe. *ASCMO* 9(1).
- Cooley, D. and S. R. Sain (2010). Spatial hierarchical modeling of precepitation extremes from a regional climate model. *Journal of Agricultural, biological and environmental statistics* 15, 381–402.
- Davison, A. C. (2003). *Statistical Models*. Cambridge University Press.
- Fisher, R. A. and L. H. C. Tippett (1928). On the estimation of the frequency distributions of the largest or smallest member of a sample. *Proceedings of the Cambridge Philosophical Society* 24, 180–190.
- Gnedenko, B. (1943). Sur la distribution limite du terme maximum d'une serie aleatoire. *Annals of mathematics*, 423–453.
- Jenkinson, A. F. (1955). The frequency distribution of the annual maximum (or minimum) values of meteorological elements. *Quarterly Journal of the Royal Meteorological Society* 81(348), 158–171.
- Wood, S. N. (2010). Fast stable restricted maximum likelihood and marginal likelihood estimation of semiparametric generalized linear models. *Journal of the Royal Statistical Society Series B: Statistical Methodology*.