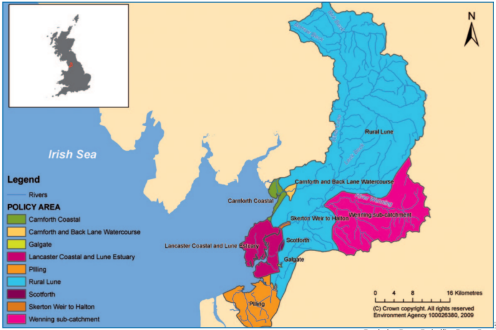
Joint estimation of extreme precipitation aggregates at different spatial scales through mixture modelling and conditional methods

Jordan Richards, Jonathan A. Tawn, Simon Brown

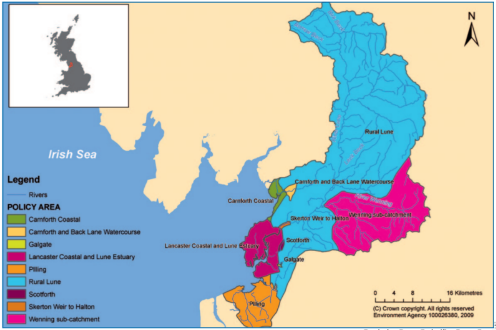
Motivation

- Extreme rainfall can potentially lead to river flooding
- Flood damage in the UK costs an estimated £1.1 billion/yr
- Storm Desmond, 2015, pictured

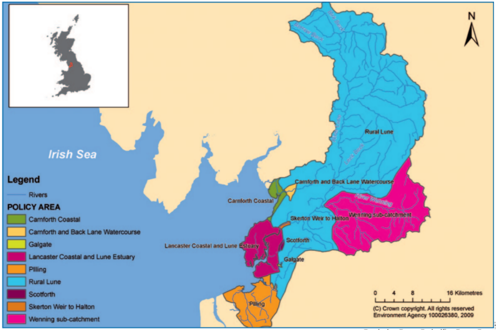
- A river breaking its banks is not typically due to intense rainfall at a single location in space
- For better flood risk, need understanding of extremal behaviour of rainfall at different spatial scales
- River Lune catchment area: $11{,}000km^2$ - If data available for inference higher spatial resolution, i.e., climate model outputs, need to aggregate up



- A river breaking its banks is not typically due to intense rainfall at a single location in space
- For better flood risk, need understanding of extremal behaviour of rainfall at different spatial scales
- River Lune catchment area: 11,000km^2 - If data available for inference higher spatial resolution, i.e., climate model outputs, need to aggregate up



- A river breaking its banks is not typically due to intense rainfall at a single location in space
- For better flood risk, need understanding of extremal behaviour of rainfall at different spatial scales
- River Lune catchment area: 11,000km^2 - If data available for inference higher spatial resolution, i.e., climate model outputs, need to aggregate up



- For spatial process $\{Y_t(s) : s \in \mathcal{S}\}$ observed at sampling locations $\mathbf{s} = (s_1, \ldots, s_d)$ and times t - for our application, $\mathbf{s}$ and t are grid-boxes
- Assume that Y_t is stationary over time, i.e., $Y_t(s) = Y(s)$ (for the moment)
- Interested in behaviour of aggregate $\bar{R}_{\mathcal{A}} = \frac{1}{|\mathcal{A}|}\int_{\mathcal{A}} Y(s)\mathrm{d}s$ for $\mathcal{A} \subset \mathcal{S}$ (or equivalent sum)
- For parsimony, we want a model that can be used for joint inference on $\bar{R}_{\mathcal{A}}$ for multiple $\mathcal{A}$ with different sizes - Pool information across different spatial scales

- For spatial process $\{Y_t(s) : s \in \mathcal{S}\}$ observed at sampling locations $\mathbf{s} = (s_1, \ldots, s_d)$ and times t - for our application, $\mathbf{s}$ and t are grid-boxes
- Assume that Y_t is stationary over time, i.e., $Y_t(s) = Y(s)$ (for the moment)
- Interested in behaviour of aggregate $\bar{R}_{\mathcal{A}} = \frac{1}{|\mathcal{A}|}\int_{\mathcal{A}} Y(s)\mathrm{d}s$ for $\mathcal{A} \subset \mathcal{S}$ (or equivalent sum)
- For parsimony, we want a model that can be used for joint inference on $\bar{R}_{\mathcal{A}}$ for multiple $\mathcal{A}$ with different sizes - Pool information across different spatial scales

- For spatial process $\{Y_t(s) : s \in \mathcal{S}\}$ observed at sampling locations $\mathbf{s} = (s_1, \ldots, s_d)$ and times t - for our application, $\mathbf{s}$ and t are grid-boxes
- Assume that Y_t is stationary over time, i.e., $Y_t(s) = Y(s)$ (for the moment)
- Interested in behaviour of aggregate $\bar{R}_{\mathcal{A}} = \frac{1}{|\mathcal{A}|} \int_{\mathcal{A}} Y(s) \mathrm{d}s$ for $\mathcal{A} \subset \mathcal{S}$ (or equivalent sum)
- For parsimony, we want a model that can be used for joint inference on $\bar{R}_{\mathcal{A}}$ for multiple $\mathcal{A}$ with different sizes - Pool information across different spatial scales

- For spatial process $\{Y_t(s) : s \in \mathcal{S}\}$ observed at sampling locations $\mathbf{s} = (s_1, \ldots, s_d)$ and times t - for our application, $\mathbf{s}$ and t are grid-boxes
- Assume that Y_t is stationary over time, i.e., $Y_t(s) = Y(s)$ (for the moment)
- Interested in behaviour of aggregate $\bar{R}_{\mathcal{A}} = \frac{1}{|\mathcal{A}|}\int_{\mathcal{A}} Y(s)\mathrm{d}s$ for $\mathcal{A} \subset \mathcal{S}$ (or equivalent sum)
- For parsimony, we want a model that can be used for joint inference on $\bar{R}_{\mathcal{A}}$ for multiple $\mathcal{A}$ with different sizes - Pool information across different spatial scales

- Particularly interested in extremal behaviour of $\bar{R}_{\mathcal{A}}$- largest events most likely to cause flooding
- Extremes are often quantified through return-levels - An $n-$year return level is a quantile expected to be exceeded on average once every n years
- Flood defences are often built within specification to withstand an $n-$year event, e.g., 100 - Storm Desmond is consider a one-in-100 year event
- Often n is greater than observation period
- Return levels can be estimated using univariate extreme value methods

- Particularly interested in extremal behaviour of $\bar{R}_{\mathcal{A}}$- largest events most likely to cause flooding
- Extremes are often quantified through return-levels - An $n-$year return level is a quantile expected to be exceeded on average once every n years
- Flood defences are often built within specification to withstand an $n-$year event, e.g., 100 - Storm Desmond is consider a one-in-100 year event
- Often n is greater than observation period
- Return levels can be estimated using univariate extreme value methods

- Particularly interested in extremal behaviour of $\bar{R}_{\mathcal{A}}$- largest events most likely to cause flooding
- Extremes are often quantified through return-levels - An $n-$year return level is a quantile expected to be exceeded on average once every n years
- Flood defences are often built within specification to withstand an $n-$year event, e.g., 100 - Storm Desmond is consider a one-in-100 year event
- Often n is greater than observation period
- Return levels can be estimated using univariate extreme value methods

- Particularly interested in extremal behaviour of $\bar{R}_{\mathcal{A}}$- largest events most likely to cause flooding
- Extremes are often quantified through return-levels - An $n-$year return level is a quantile expected to be exceeded on average once every n years
- Flood defences are often built within specification to withstand an $n-$year event, e.g., 100 - Storm Desmond is consider a one-in-100 year event
- Often n is greater than observation period
- Return levels can be estimated using univariate extreme value methods

- Particularly interested in extremal behaviour of $\bar{R}_{\mathcal{A}}$- largest events most likely to cause flooding
- Extremes are often quantified through return-levels - An $n-$year return level is a quantile expected to be exceeded on average once every n years
- Flood defences are often built within specification to withstand an $n-$year event, e.g., 100 - Storm Desmond is consider a one-in-100 year event
- Often n is greater than observation period
- Return levels can be estimated using univariate extreme value methods

GPD modelling overview

For univariate r.v. Y, we assume that there exists a high threshold $u > 0$:

- $(Y - u)|Y > u \sim \text{GPD}(\sigma, \xi)$ for $\sigma > 0, \xi \in \mathbb{R}$
- Estimate parameters and then re-scale to find $\Pr(Y > y)$. Often estimate $\Pr(Y > u)$ empirically
- The $n-$year return level for Y is the $1 - 1/(n \times r_y)$ quantile of Y, for r_y observations per year

GPD modelling overview

For univariate r.v. Y, we assume that there exists a high threshold $u > 0$:

- $(Y - u)|Y > u \sim \text{GPD}(\sigma, \xi)$ for $\sigma > 0, \xi \in \mathbb{R}$
- Estimate parameters and then re-scale to find $\Pr(Y > y)$. Often estimate $\Pr(Y > u)$ empirically
- The $n-$year return level for Y is the $1 - 1/(n \times r_y)$ quantile of Y, for r_y observations per year

GPD modelling overview

For univariate r.v. Y, we assume that there exists a high threshold $u > 0$:

- $(Y - u)|Y > u \sim \text{GPD}(\sigma, \xi)$ for $\sigma > 0, \xi \in \mathbb{R}$
- Estimate parameters and then re-scale to find $\Pr(Y > y)$. Often estimate $\Pr(Y > u)$ empirically
- The $n-$year return level for Y is the $1 - 1/(n \times r_y)$ quantile of Y, for r_y observations per year

Modelling $\bar{R}_{\mathcal{A}}$

- Could simply take sample aggregates $\bar{R}_{\mathcal{A}}$ and fit GPD model
- Wasteful - Less data for modelling and lose all information from the underlying process, e.g., margins, dependence
- No pooling of info across different scales as separate fits for each $\mathcal{A}$

Modelling $\bar{R}_{\mathcal{A}}$

- Could simply take sample aggregates $\bar{R}_{\mathcal{A}}$ and fit GPD model
- Wasteful - Less data for modelling and lose all information from the underlying process, e.g., margins, dependence
- No pooling of info across different scales as separate fits for each $\mathcal{A}$

Modelling $\bar{R}_{\mathcal{A}}$

- Could simply take sample aggregates $\bar{R}_{\mathcal{A}}$ and fit GPD model
- Wasteful - Less data for modelling and lose all information from the underlying process, e.g., margins, dependence
- No pooling of info across different scales as separate fits for each $\mathcal{A}$

Proposed method

- We model the high-resolution process $\{Y(s) : s \in \mathcal{S}\}$
- Simulate fields from the fitted model over entire spatial domain $\mathcal{S}$
- Derive $\bar{R}_{\mathcal{A}}$ from simulated fields and for different $\mathcal{A}$
- Use simulated samples of $\bar{R}_{\mathcal{A}}$ for inference, i.e., parametric "bootstrap" for $\bar{R}_{\mathcal{A}}$ using $Y(s)$

Proposed method

- We model the high-resolution process $\{Y(s) : s \in \mathcal{S}\}$
- Simulate fields from the fitted model over entire spatial domain $\mathcal{S}$
- Derive $\bar{R}_{\mathcal{A}}$ from simulated fields and for different $\mathcal{A}$
- Use simulated samples of $\bar{R}_{\mathcal{A}}$ for inference, i.e., parametric "bootstrap" for $\bar{R}_{\mathcal{A}}$ using $Y(s)$

Proposed method

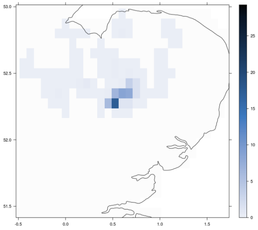
- We model the high-resolution process $\{Y(s) : s \in \mathcal{S}\}$
- Simulate fields from the fitted model over entire spatial domain $\mathcal{S}$
- Derive $\bar{R}_{\mathcal{A}}$ from simulated fields and for different $\mathcal{A}$
- Use simulated samples of $\bar{R}_{\mathcal{A}}$ for inference, i.e., parametric "bootstrap" for $\bar{R}_{\mathcal{A}}$ using $Y(s)$

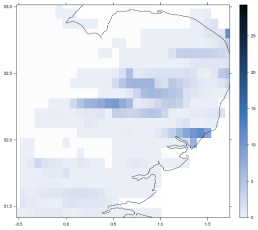
Proposed method

- We model the high-resolution process $\{Y(s) : s \in \mathcal{S}\}$
- Simulate fields from the fitted model over entire spatial domain $\mathcal{S}$
- Derive $\bar{R}_{\mathcal{A}}$ from simulated fields and for different $\mathcal{A}$
- Use simulated samples of $\bar{R}_{\mathcal{A}}$ for inference, i.e., parametric “bootstrap" for $\bar{R}_{\mathcal{A}}$ using $Y(s)$

Data

- From UKCP18 - Hourly precipitation rate (mm/hour), Summer (JJA), 1980-2000
- Values assigned to $d = 934$ spatial grid-boxes rather than point locations - Land only
- Convection permitting model - Spatial resolution $(5km)^2$





Model $Y(s)$ overview

- Model margins and dependence separately
- Focus on modelling extremes of process - largest values of underlying process produce largest values of aggregate
- Procedure - common practice for spatial extremes modelling -
  - Fit a marginal model at each site $s \in \mathcal{S}$
  - Standardise observed data to $\{X(s) : s \in \mathcal{S}\}$ - Fit extremal dependence model to $X(s)$

Model $Y(s)$ overview

- Model margins and dependence separately
- Focus on modelling extremes of process - largest values of underlying process produce largest values of aggregate
- Procedure - common practice for spatial extremes modelling -
  - Fit a marginal model at each site $s \in \mathcal{S}$
  - Standardise observed data to $\{X(s) : s \in \mathcal{S}\}$ - Fit extremal dependence model to $X(s)$

Model $Y(s)$ overview

- Model margins and dependence separately
- Focus on modelling extremes of process - largest values of underlying process produce largest values of aggregate
- Procedure - common practice for spatial extremes modelling -
  - Fit a marginal model at each site $s \in \mathcal{S}$
  - Standardise observed data to $\{X(s) : s \in \mathcal{S}\}$ - Fit extremal dependence model to $X(s)$

Model $Y(s)$ overview

- Model margins and dependence separately
- Focus on modelling extremes of process - largest values of underlying process produce largest values of aggregate
- Procedure - common practice for spatial extremes modelling -
  - Fit a marginal model at each site $s \in \mathcal{S}$
  - Standardise observed data to $\{X(s) : s \in \mathcal{S}\}$ - Fit extremal dependence model to $X(s)$

Model $Y(s)$ overview

- Model margins and dependence separately
- Focus on modelling extremes of process - largest values of underlying process produce largest values of aggregate
- Procedure - common practice for spatial extremes modelling -
  - Fit a marginal model at each site $s \in \mathcal{S}$
  - Standardise observed data to $\{X(s) : s \in \mathcal{S}\}$ - Fit extremal dependence model to $X(s)$

Marginal model

The marginal model $F_s(y)$ has three components:

- If $Y(s) = 0$, then $F_s(y) = p(s)$, i.e., probability of no rain - This is used as a censoring threshold in inference and simulation
- If $0 < Y(s) < q(s)$, then $F_s(y)$ is estimated empirically - Here $q(s)$ is the $\lambda-$quantile of $Y(s)$
- If $Y(s) \geq q(s)$, then $Y(s) - q(s) \sim \text{GPD}(\sigma(s), \xi(s))$

The four parameters $p(s), q(s), \sigma(s), \xi(s)$ are represented as functions of location s and elevation using Generalised Additive Models (GAMs) [Wood, 2006, Youngman, 2019]

Marginal model

The marginal model $F_s(y)$ has three components:

- If $Y(s) = 0$, then $F_s(y) = p(s)$, i.e., probability of no rain - This is used as a censoring threshold in inference and simulation
- If $0 < Y(s) < q(s)$, then $F_s(y)$ is estimated empirically - Here $q(s)$ is the $\lambda-$quantile of $Y(s)$
- If $Y(s) \geq q(s)$, then $Y(s) - q(s) \sim \text{GPD}(\sigma(s), \xi(s))$

The four parameters $p(s), q(s), \sigma(s), \xi(s)$ are represented as functions of location s and elevation using Generalised Additive Models (GAMs) [Wood, 2006, Youngman, 2019]

Marginal model

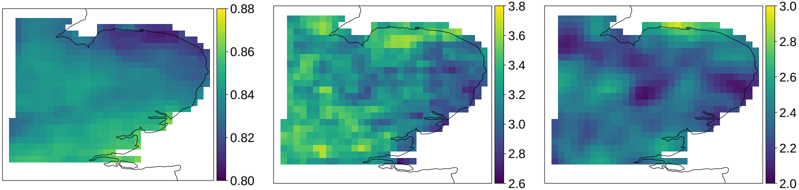
The marginal model $F_s(y)$ has three components:

- If $Y(s) = 0$, then $F_s(y) = p(s)$, i.e., probability of no rain - This is used as a censoring threshold in inference and simulation
- If $0 < Y(s) < q(s)$, then $F_s(y)$ is estimated empirically - Here $q(s)$ is the $\lambda-$quantile of $Y(s)$
- If $Y(s) \geq q(s)$, then $Y(s) - q(s) \sim \text{GPD}(\sigma(s), \xi(s))$

The four parameters $p(s), q(s), \sigma(s), \xi(s)$ are represented as functions of location s and elevation using Generalised Additive Models (GAMs) [Wood, 2006, Youngman, 2019]

Just for visualisation:

- Left: $p(s)$. Middle: $q(s)$ for $\lambda = 0.995$. Right: $\sigma(s)$.



The marginal fits are used to standardise the data to fit the extremal dependence model, a spatial extension of the m.v. conditional extremes framework [Heffernan and Tawn, 2004].

Multivariate conditional extremes (simplified)

For a random $d-$vector $\mathbf{X}$ with standard exponential upper-tailed margins:

- Conditioning on $X_i > u$ for $u > 0$ and any $i = 1, \ldots, d$
- For normalising vectors $\alpha \in [0,1]^{d-1}$ and $\beta \in [0,1]^{d-1}$
- Assume that

$$\left(\frac{\mathbf{X}_{-i} - \alpha X_i}{X_i^{\beta}}, X_i - u\right) \Big| \left(X_i > u\right) \xrightarrow{d} \mathbf{Z}E,$$

  as $u \to \infty$ and where $E \sim \text{Exp}(1)$ independent of $\mathbf{Z}$, operations are taken componentwise and the residual dist. of $\mathbf{Z}$ has non-degenerate marginals.
- Make some assumptions about $\mathbf{Z}$ to model $\mathbf{X}_{-i} | X_i > u$

Multivariate conditional extremes (simplified)

For a random $d-$vector $\mathbf{X}$ with standard exponential upper-tailed margins:

- Conditioning on $X_i > u$ for $u > 0$ and any $i = 1, \ldots, d$
- For normalising vectors $\boldsymbol{\alpha} \in [0,1]^{d-1}$ and $\boldsymbol{\beta} \in [0,1]^{d-1}$
- Assume that

$$\left(\frac{\mathbf{X}_{-i} - \boldsymbol{\alpha} X_i}{X_i^{\boldsymbol{\beta}}}, X_i - u \right) \Bigg| \left(X_i > u \right) \xrightarrow{d} \mathbf{Z} E,$$

as $u \to \infty$ and where $E \sim \text{Exp}(1)$ independent of $\mathbf{Z}$, operations are taken componentwise and the residual dist. of $\mathbf{Z}$ has non-degenerate marginals.

- Make some assumptions about $\mathbf{Z}$ to model $\mathbf{X}_{-i} | X_i > u$

Multivariate conditional extremes (simplified)

For a random $d-$vector $\mathbf{X}$ with standard exponential upper-tailed margins:

- Conditioning on $X_i > u$ for $u > 0$ and any $i = 1, \ldots, d$
- For normalising vectors $\boldsymbol{\alpha} \in [0,1]^{d-1}$ and $\boldsymbol{\beta} \in [0,1]^{d-1}$
- Assume that

$$\left(\frac{\mathbf{X}_{-i} - \boldsymbol{\alpha} X_i}{X_i^{\boldsymbol{\beta}}}, X_i - u \right) \Bigg| \left(X_i > u \right) \xrightarrow{d} \mathbf{Z} E,$$

  as $u \to \infty$ and where $E \sim \text{Exp}(1)$ independent of $\mathbf{Z}$, operations are taken componentwise and the residual dist. of $\mathbf{Z}$ has non-degenerate marginals.

- Make some assumptions about $\mathbf{Z}$ to model $\mathbf{X}_{-i} | X_i > u$

Multivariate conditional extremes (simplified)

For a random $d-$vector $\mathbf{X}$ with standard exponential upper-tailed margins:

- Conditioning on $X_i > u$ for $u > 0$ and any $i = 1, \ldots, d$
- For normalising vectors $\boldsymbol{\alpha} \in [0,1]^{d-1}$ and $\boldsymbol{\beta} \in [0,1]^{d-1}$
- Assume that

$$\left(\frac{\mathbf{X}_{-i} - \boldsymbol{\alpha} X_i}{X_i^{\boldsymbol{\beta}}}, X_i - u \right) \Bigg| \Big(X_i > u \Big) \xrightarrow{d} \mathbf{Z}E,$$

  as $u \to \infty$ and where $E \sim \text{Exp}(1)$ independent of $\mathbf{Z}$, operations are taken componentwise and the residual dist. of $\mathbf{Z}$ has non-degenerate marginals.
- Make some assumptions about $\mathbf{Z}$ to model $\mathbf{X}_{-i} | X_i > u$

Assuming limit holds for large u, we have the regression type model:

$$\mathbf{X}_{-i}|X_i > u = \boldsymbol{\alpha} X_i + X_i^{\boldsymbol{\beta}} \mathbf{Z}$$

- $\mathbf{Z}$ typically assumed to be non-standard Gaussian
- α values control strength of extremal dependence
- β controls the effect that X_i has on the variance of $\mathbf{X}_{-i}|X_i > u$

Assuming limit holds for large u, we have the regression type model:

$$\mathbf{X}_{-i}|X_i > u = \boldsymbol{\alpha} X_i + X_i^{\boldsymbol{\beta}} \mathbf{Z}$$

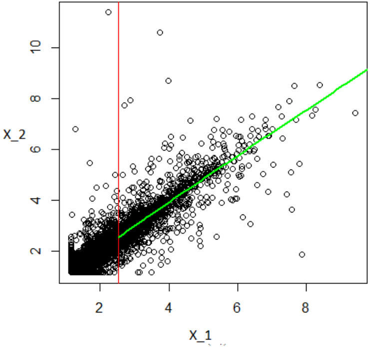
- $\mathbf{Z}$ typically assumed to be non-standard Gaussian
- α values control strength of extremal dependence
- β controls the effect that X_i has on the variance of $\mathbf{X}_{-i}|X_i > u$

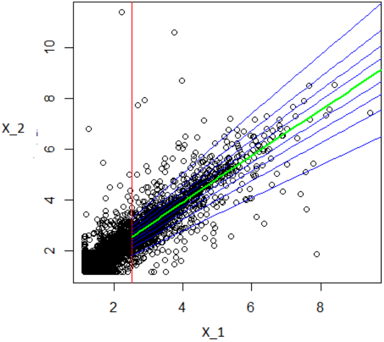
Assuming limit holds for large u, we have the regression type model:

$$\mathbf{X}_{-i}|X_i > u = \boldsymbol{\alpha} X_i + X_i^{\boldsymbol{\beta}}\mathbf{Z}$$

- $\mathbf{Z}$ typically assumed to be non-standard Gaussian
- α values control strength of extremal dependence
- β controls the effect that X_i has on the variance of $\mathbf{X}_{-i}|X_i > u$

- $X_2|(X_1 = x) = \alpha x + x^{\beta} Z, x \geq u$
- Expected value $\alpha x + x^{\beta} \times \mathbb{E}[Z]$ and variance $x^{2\beta}\text{Var}(Z)$





Effect of $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ on extremal dependence class. For α_j, β_j corresponding to the X_j component of $\mathbf{X}_{-i}$:

- If $\alpha_j = 1$ and $\beta_j = 0$, then X_j and X_i are asymptotically dependent
- Decreasing strength of extremal dependence with α_j and β_j
- If $\alpha_j < 1 \Rightarrow$ asymptotically independent and near-independence for $\alpha_j = \beta_j = 0$

Effect of $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ on extremal dependence class. For α_j, β_j corresponding to the X_j component of $\mathbf{X}_{-i}$:

- If $\alpha_j = 1$ and $\beta_j = 0$, then X_j and X_i are asymptotically dependent
- Decreasing strength of extremal dependence with α_j and β_j
- If $\alpha_j < 1 \Rightarrow$ asymptotically independent and near-independence for $\alpha_j = \beta_j = 0$

Effect of $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ on extremal dependence class. For α_j, β_j corresponding to the X_j component of $\mathbf{X}_{-i}$:

- If $\alpha_j = 1$ and $\beta_j = 0$, then X_j and X_i are asymptotically dependent
- Decreasing strength of extremal dependence with α_j and β_j
- If $\alpha_j < 1 \Rightarrow$ asymptotically independent and near-independence for $\alpha_j = \beta_j = 0$

Extremal dependence

Extremal dependence can be quantified through the measure

$$\chi = \lim_{p \to 1} \Pr\{F_1(X_1) > p | F_2(X_2) > p\}.$$

- If $\chi > 0$, then X_1, X_2 are asymptotically dependent - $\chi = 0 \Rightarrow$ asymptotic independent
- Classical spatial extremes models can only handle one class, e.g., max-stable/Pareto process are AD only, Gaussian processes are AI only
- [Heffernan and Tawn, 2004] can accommodate both classes

Extremal dependence

Extremal dependence can be quantified through the measure

$$\chi = \lim_{p \to 1} \Pr\{F_1(X_1) > p | F_2(X_2) > p\}.$$

- If $\chi > 0$, then X_1, X_2 are asymptotically dependent - $\chi = 0 \Rightarrow$ asymptotic independent
- Classical spatial extremes models can only handle one class, e.g., max-stable/Pareto process are AD only, Gaussian processes are AI only
- [Heffernan and Tawn, 2004] can accommodate both classes

Extremal dependence

Extremal dependence can be quantified through the measure

$$\chi = \lim_{p \to 1} \Pr\{F_1(X_1) > p | F_2(X_2) > p\}.$$

- If $\chi > 0$, then X_1, X_2 are asymptotically dependent - $\chi = 0 \Rightarrow$ asymptotic independent
- Classical spatial extremes models can only handle one class, e.g., max-stable/Pareto process are AD only, Gaussian processes are AI only
- [Heffernan and Tawn, 2004] can accommodate both classes

Spatial dependence model - general outline

General outline for spatial conditional extremes:
[Wadsworth and Tawn, 2019]

- Spatial extension of [Heffernan and Tawn, 2004] multivariate conditional extremes
- Spatial process $\{X(s) : s \in \mathcal{S}\}$ with standard exponential upper-tailed margins (standard Laplace)
- Condition on high value of process at conditioning site $s_O \in \mathcal{S}$, i.e., $X(s_O) > u$
- Model (normalised) process

More formally...

Spatial dependence model - general outline

General outline for spatial conditional extremes:
[Wadsworth and Tawn, 2019]

- Spatial extension of [Heffernan and Tawn, 2004] multivariate conditional extremes
- Spatial process $\{X(s) : s \in \mathcal{S}\}$ with standard exponential upper-tailed margins (standard Laplace)
- Condition on high value of process at conditioning site $s_O \in \mathcal{S}$, i.e., $X(s_O) > u$
- Model (normalised) process

More formally...

Spatial dependence model - general outline

General outline for spatial conditional extremes:
[Wadsworth and Tawn, 2019]

- Spatial extension of [Heffernan and Tawn, 2004] multivariate conditional extremes
- Spatial process $\{X(s) : s \in \mathcal{S}\}$ with standard exponential upper-tailed margins (standard Laplace)
- Condition on high value of process at conditioning site $s_O \in \mathcal{S}$, i.e., $X(s_O) > u$
- Model (normalised) process

More formally...

Spatial dependence model - general outline

General outline for spatial conditional extremes:
[Wadsworth and Tawn, 2019]

- Spatial extension of [Heffernan and Tawn, 2004] multivariate conditional extremes
- Spatial process $\{X(s) : s \in \mathcal{S}\}$ with standard exponential upper-tailed margins (standard Laplace)
- Condition on high value of process at conditioning site $s_O \in \mathcal{S}$, i.e., $X(s_O) > u$
- Model (normalised) process

More formally...

Conditional spatial extremes

Assume there exists normalising functions $\alpha : [0, \infty) \to [0, 1]$, with $\alpha(0) = 1$, and $\beta : [0, \infty) \to [0, 1]$, such that

$$\left(\left\{ \frac{X(s) - \alpha\{h(s, s_O)\}X(s_O)}{\{X(s_O)\}^{\beta\{h(s,s_O)\}}} : s \in \mathcal{S} \right\}, X(s_O) - u \right) \Bigg| \Big(X(s_O) > u \Big)$$

$$\xrightarrow{d} \left(\Big\{ Z(s|s_O) : s \in \mathcal{S} \Big\}, E \right) \text{ as } u \to \infty,$$

- (Residual) process $\{Z(s|s_O) : s \in \mathcal{S}\}$ independent of $E \sim \text{Exp}(1)$, satisfies $Z(s_O|s_O) = 0$ almost surely
- $h(s, s_O)$ anisotropic distance between s and s_O
- Stationarity assumption; same relationships for any s_O

Modelling choices - Normalising functions

- Positive extremal dependence between $X(s)$ and $X(s_O)$ decreases with α and β for $s \in \mathcal{S}$
- We let α and β decrease with distance $h(s, s_O)$, i.e., decreasing extremal dependence
- If $\alpha\{h(s, s_O)\} = 1$ and $\beta\{h(s, s_O)\} = 0$, then $X(s)$ and $X(s_O)$ are asymptotically dependent
- $$\alpha(h) = \begin{cases} 1, & h \leq \Delta, \\ \exp(-\{(h-\Delta)/\kappa_{\alpha_1}\}^{\kappa_{\alpha_2}}), & h > \Delta, \end{cases}$$

  where $\Delta \geq 0$ and $\kappa_{\alpha_1}, \kappa_{\alpha_2} > 0$.

Modelling choices - Normalising functions

- Positive extremal dependence between $X(s)$ and $X(s_O)$ decreases with α and β for $s \in \mathcal{S}$
- We let α and β decrease with distance $h(s, s_O)$, i.e., decreasing extremal dependence
- If $\alpha\{h(s, s_O)\} = 1$ and $\beta\{h(s, s_O)\} = 0$, then $X(s)$ and $X(s_O)$ are asymptotically dependent
- $$\alpha(h) = \begin{cases} 1, & h \leq \Delta, \\ \exp(-\{(h - \Delta)/\kappa_{\alpha_1}\}^{\kappa_{\alpha_2}}), & h > \Delta, \end{cases}$$

  where $\Delta \geq 0$ and $\kappa_{\alpha_1}, \kappa_{\alpha_2} > 0$.

Modelling choices - Normalising functions

- Positive extremal dependence between $X(s)$ and $X(s_O)$ decreases with α and β for $s \in \mathcal{S}$
- We let α and β decrease with distance $h(s, s_O)$, i.e., decreasing extremal dependence
- If $\alpha\{h(s, s_O)\} = 1$ and $\beta\{h(s, s_O)\} = 0$, then $X(s)$ and $X(s_O)$ are asymptotically dependent
- $$\alpha(h) = \begin{cases} 1, & h \leq \Delta, \\ \exp(-\{(h-\Delta)/\kappa_{\alpha_1}\}^{\kappa_{\alpha_2}}), & h > \Delta, \end{cases}$$
  where $\Delta \geq 0$ and $\kappa_{\alpha_1}, \kappa_{\alpha_2} > 0$.

Modelling choices - Normalising functions

- Positive extremal dependence between $X(s)$ and $X(s_O)$ decreases with α and β for $s \in \mathcal{S}$
- We let α and β decrease with distance $h(s, s_O)$, i.e., decreasing extremal dependence
- If $\alpha\{h(s, s_O)\} = 1$ and $\beta\{h(s, s_O)\} = 0$, then $X(s)$ and $X(s_O)$ are asymptotically dependent
- $$\alpha(h) = \begin{cases} 1, & h \leq \Delta, \\ \exp(-\{(h - \Delta)/\kappa_{\alpha_1}\}^{\kappa_{\alpha_2}}), & h > \Delta, \end{cases}$$

  where $\Delta \geq 0$ and $\kappa_{\alpha_1}, \kappa_{\alpha_2} > 0$.

Residual process construction - Dependence

Modelling choice for dependence in $\{Z(s|s_O)\}$:

- Start with a GP $\{W(s) : s \in \mathcal{S}\}$ with a stationary correlation structure
- Set $\{W(s|s_O) : s \in \mathcal{S}\} = \{W(s)|W(s_O) = 0 : s \in \mathcal{S}\}$
- Marginal change to $\{Z(s|s_O)\}$

Residual process - Margins

Modelling choice for margins of $\{Z(s|s_O) : s \in \mathcal{S}\}$:

- Set margins to delta-Laplace

$$f(z) = \frac{\delta}{2k\sigma\Gamma\left(\frac{1}{\delta}\right)} \exp\left\{-\left|\frac{z-\mu}{k\sigma}\right|^{\delta}\right\},$$

  with $\Gamma(\cdot)$ as the standard gamma function and $\mu \in \mathbb{R}, \sigma > 0, \delta > 0$ and $k^2 = \Gamma(1/\delta)\Gamma(3/\delta)$.
- More flexible than Laplace ($\delta = 1$) or Gaussian ($\delta = 2$)
- Parametrise μ, σ and δ as functions of $h(s, s_O)$ - For $Z(s_O|s_O) = 0$ almost surely, require $\mu(0) = \sigma(0) = 0$.

Inference

- Censoring makes using full likelihood inference computationally infeasible
- Use a triple-wise pseudo-censored-likelihood for inference, i.e., conditioning site and two other locations
- Only sites within a specified maximum distance - We use around $25km$, can be chosen through validation techniques
- Use only a sub-sample of possible triples - Good fit using only 5000 triples (of a possible ≥ 400 million!)
- Model can be fit to high-dimensional data - $d = 934$ with $d > 7000$ in extensions

Inference

- Censoring makes using full likelihood inference computationally infeasible
- Use a triple-wise pseudo-censored-likelihood for inference, i.e., conditioning site and two other locations
- Only sites within a specified maximum distance - We use around 25km, can be chosen through validation techniques
- Use only a sub-sample of possible triples - Good fit using only 5000 triples (of a possible $\geq$ 400 million!)
- Model can be fit to high-dimensional data - $d = 934$ with $d > 7000$ in extensions

Inference

- Censoring makes using full likelihood inference computationally infeasible
- Use a triple-wise pseudo-censored-likelihood for inference, i.e., conditioning site and two other locations
- Only sites within a specified maximum distance - We use around 25*km*, can be chosen through validation techniques
- Use only a sub-sample of possible triples - Good fit using only 5000 triples (of a possible $\geq$ 400 million!)
- Model can be fit to high-dimensional data - $d = 934$ with $d > 7000$ in extensions

Inference

- Censoring makes using full likelihood inference computationally infeasible
- Use a triple-wise pseudo-censored-likelihood for inference, i.e., conditioning site and two other locations
- Only sites within a specified maximum distance - We use around 25km, can be chosen through validation techniques
- Use only a sub-sample of possible triples - Good fit using only 5000 triples (of a possible $\geq$ 400 million!)
- Model can be fit to high-dimensional data - $d = 934$ with $d > 7000$ in extensions

Inference

- Censoring makes using full likelihood inference computationally infeasible
- Use a triple-wise pseudo-censored-likelihood for inference, i.e., conditioning site and two other locations
- Only sites within a specified maximum distance - We use around $25km$, can be chosen through validation techniques
- Use only a sub-sample of possible triples - Good fit using only 5000 triples (of a possible $\geq$ 400 million!)
- Model can be fit to high-dimensional data - $d = 934$ with $d > 7000$ in extensions

Simulation

- Choose a conditioning site s_O uniformly at random
- Simulate $X(s_O) > u = u + \text{Exp}(1)$
- Simulate $\{X(s)|X(s_O) > u\}$ for all $s \in \mathcal{S}$
- Through repeated sampling and importance sampling, get approximate sample from $\{X(s)|\max_{s\in\mathcal{S}} X(s) > u\}$
- Use observations for $\{X(s)|\max_{s\in\mathcal{S}} X(s) < u\}$ to get unconditioned $\{X(s)\}$
- Transform $\{X(s)\}$ to $\{Y(s)\}$

Simulation

- Choose a conditioning site s_O uniformly at random
- Simulate $X(s_O) > u = u + \text{Exp}(1)$
- Simulate $\{X(s)|X(s_O) > u\}$ for all $s \in \mathcal{S}$
- Through repeated sampling and importance sampling, get approximate sample from $\{X(s)|\max_{s\in\mathcal{S}} X(s) > u\}$
- Use observations for $\{X(s)|\max_{s\in\mathcal{S}} X(s) < u\}$ to get unconditioned $\{X(s)\}$
- Transform $\{X(s)\}$ to $\{Y(s)\}$

Simulation

- Choose a conditioning site s_O uniformly at random
- Simulate $X(s_O) > u = u + \text{Exp}(1)$
- Simulate $\{X(s)|X(s_O) > u\}$ for all $s \in \mathcal{S}$
- Through repeated sampling and importance sampling, get approximate sample from $\{X(s)|\max_{s\in\mathcal{S}} X(s) > u\}$
- Use observations for $\{X(s)|\max_{s\in\mathcal{S}} X(s) < u\}$ to get unconditioned $\{X(s)\}$
- Transform $\{X(s)\}$ to $\{Y(s)\}$

Simulation

- Choose a conditioning site s_O uniformly at random
- Simulate $X(s_O) > u = u + \text{Exp}(1)$
- Simulate $\{X(s)|X(s_O) > u\}$ for all $s \in \mathcal{S}$
- Through repeated sampling and importance sampling, get approximate sample from $\{X(s)|\max_{s\in\mathcal{S}} X(s) > u\}$
- Use observations for $\{X(s)|\max_{s\in\mathcal{S}} X(s) < u\}$ to get unconditioned $\{X(s)\}$
- Transform $\{X(s)\}$ to $\{Y(s)\}$

Simulation

- Choose a conditioning site s_O uniformly at random
- Simulate $X(s_O) > u = u + \text{Exp}(1)$
- Simulate $\{X(s)|X(s_O) > u\}$ for all $s \in \mathcal{S}$
- Through repeated sampling and importance sampling, get approximate sample from $\{X(s)|\max_{s\in\mathcal{S}} X(s) > u\}$
- Use observations for $\{X(s)|\max_{s\in\mathcal{S}} X(s) < u\}$ to get unconditioned $\{X(s)\}$
- Transform $\{X(s)\}$ to $\{Y(s)\}$

Simulation

- Choose a conditioning site s_O uniformly at random
- Simulate $X(s_O) > u = u + \text{Exp}(1)$
- Simulate $\{X(s)|X(s_O) > u\}$ for all $s \in \mathcal{S}$
- Through repeated sampling and importance sampling, get approximate sample from $\{X(s)| \max_{s\in\mathcal{S}} X(s) > u\}$
- Use observations for $\{X(s)| \max_{s\in\mathcal{S}} X(s) < u\}$ to get unconditioned $\{X(s)\}$
- Transform $\{X(s)\}$ to $\{Y(s)\}$

Dependence Model - Parametric function choice

To get an idea of the functional forms of the dependence parameters, we fit a simple dependence model:

- No dependence in residual process $\{Z(s|s_O) : s \in \mathcal{S}\}$
- Individual parameter estimates, rather than fitted functions i.e., sequence of α_{s_i} for $i = 1, \ldots, d$, not $\alpha\{h(s, s_O)\}$
- Done for several conditioning sites spread out over domain

Dependence Model - Parametric function choice

To get an idea of the functional forms of the dependence parameters, we fit a simple dependence model:

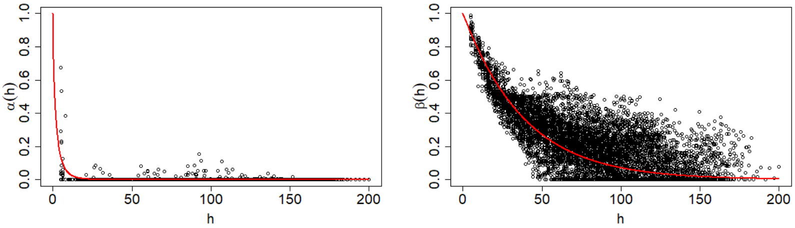
- No dependence in residual process $\{Z(s|s_O) : s \in \mathcal{S}\}$
- Individual parameter estimates, rather than fitted functions i.e., sequence of α_{s_i} for $i = 1, \ldots, d$, not $\alpha\{h(s, s_O)\}$
- Done for several conditioning sites spread out over domain

Dependence Model - Parametric function choice

To get an idea of the functional forms of the dependence parameters, we fit a simple dependence model:

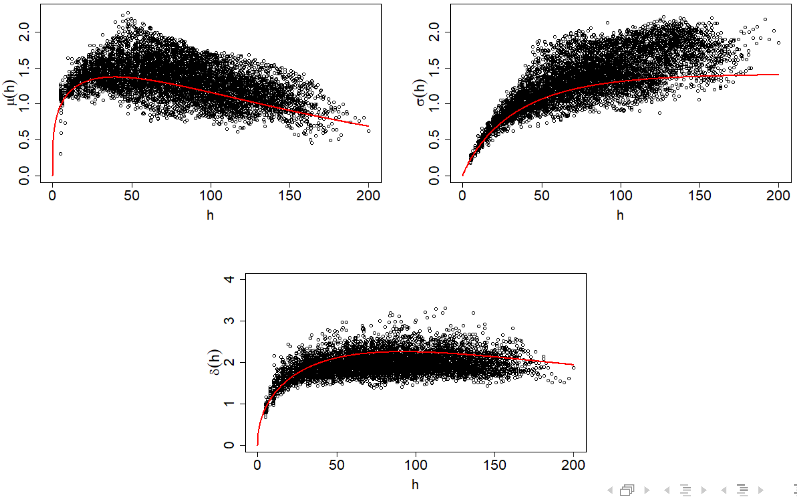
- No dependence in residual process $\{Z(s|s_O) : s \in \mathcal{S}\}$
- Individual parameter estimates, rather than fitted functions i.e., sequence of α_{s_i} for $i = 1, \ldots, d$, not $\alpha\{h(s, s_O)\}$
- Done for several conditioning sites spread out over domain

Alpha/Beta



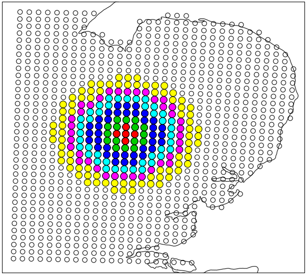
α(h)
0.0
0.2
0.4
0.6
0.8
1.0
0
50
100
150
200
h
β(h)
0.0
0.2
0.4
0.6
0.8
1.0
0
50
100
150
200
h

Mu/Sigma/Delta



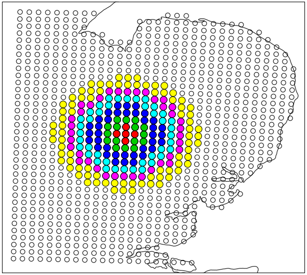
Aggregate diagnostics

- Do replications from $\{Y(s) : s \in \mathcal{S}\}$ capture tail of $\bar{R}_{\mathcal{A}}$?
- Simulate over entire domain, roughly 200 years of events
- Aggregate over increasingly large regions (coloured points and interior) $(125 - 5425) - km^2$
- Compare model quantiles against data



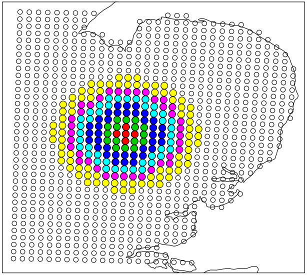
Aggregate diagnostics

- Do replications from $\{Y(s) : s \in \mathcal{S}\}$ capture tail of $\bar{R}_{\mathcal{A}}$?
- Simulate over entire domain, roughly 200 years of events
- Aggregate over increasingly large regions (coloured points and interior) $(125 - 5425) - km^2$
- Compare model quantiles against data



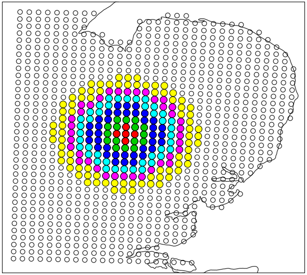
Aggregate diagnostics

- Do replications from $\{Y(s) : s \in \mathcal{S}\}$ capture tail of $\bar{R}_{\mathcal{A}}$?
- Simulate over entire domain, roughly 200 years of events
- Aggregate over increasingly large regions (coloured points and interior) $(125 - 5425) - km^2$
- Compare model quantiles against data



Aggregate diagnostics

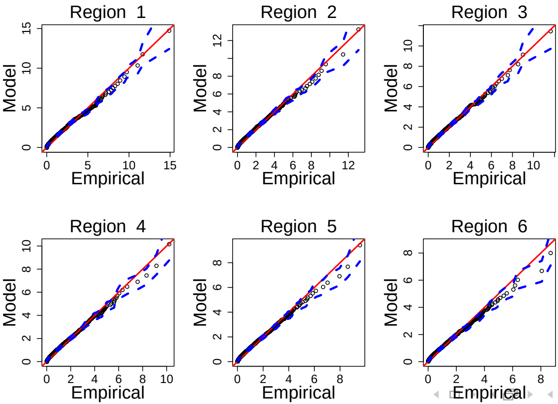
- Do replications from $\{Y(s) : s \in \mathcal{S}\}$ capture tail of $\bar{R}_{\mathcal{A}}$?
- Simulate over entire domain, roughly 200 years of events
- Aggregate over increasingly large regions (coloured points and interior) $(125 - 5425) - km^2$
- Compare model quantiles against data



Model diagnostics

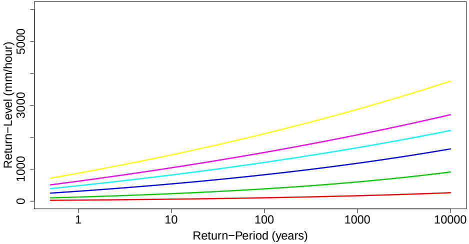
Q-Q plots (blue 95% confidence bands) for high quantiles:

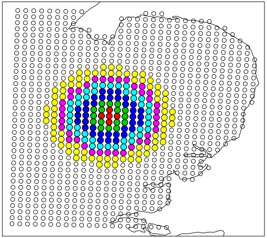
- Regions increase in size with label (i.e., 1 smallest - 6 biggest)
- Largest quantile corresponds to a 20 - year return level



Results

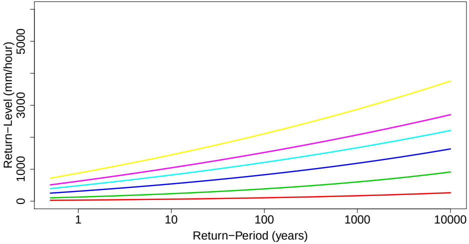
- GPD modelling for return levels of total volume of rainfall
- Guaranteed physically-consistent estimates, i.e., no overlap - not guaranteed by univariate methods

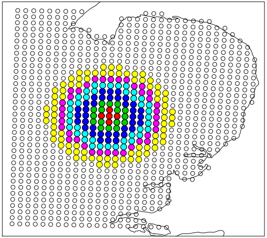




Results

- GPD modelling for return levels of total volume of rainfall
- Guaranteed physically-consistent estimates, i.e., no overlap - not guaranteed by univariate methods





Mixture modelling - Motivation

- Extreme precipitation driven by mixture of spatially localised, high intensity (convective) events and large-scale, low-intensity events (non-convective/frontal)
- In original work, we only model convective events - Recall $Y(s_O) > u$
- For small $\mathcal{A}$, large values of $\bar{R}_{\mathcal{A}}$ given by convective events
- As $|\mathcal{A}|$ increases, non-convective events become more important
- To jointly model $\bar{R}_{\mathcal{A}}$ for small and large $\mathcal{A}$, need a model for $Y(s)$ that can capture both types of events

Mixture modelling - Motivation

- Extreme precipitation driven by mixture of spatially localised, high intensity (convective) events and large-scale, low-intensity events (non-convective/frontal)
- In original work, we only model convective events - Recall $Y(s_O) > u$
- For small $\mathcal{A}$, large values of $\bar{R}_{\mathcal{A}}$ given by convective events
- As $|\mathcal{A}|$ increases, non-convective events become more important
- To jointly model $\bar{R}_{\mathcal{A}}$ for small and large $\mathcal{A}$, need a model for $Y(s)$ that can capture both types of events

Mixture modelling - Motivation

- Extreme precipitation driven by mixture of spatially localised, high intensity (convective) events and large-scale, low-intensity events (non-convective/frontal)
- In original work, we only model convective events - Recall $Y(s_O) > u$
- For small $\mathcal{A}$, large values of $\bar{R}_{\mathcal{A}}$ given by convective events
- As $|\mathcal{A}|$ increases, non-convective events become more important
- To jointly model $\bar{R}_{\mathcal{A}}$ for small and large $\mathcal{A}$, need a model for $Y(s)$ that can capture both types of events

Mixture modelling - Motivation

- Extreme precipitation driven by mixture of spatially localised, high intensity (convective) events and large-scale, low-intensity events (non-convective/frontal)
- In original work, we only model convective events - Recall $Y(s_O) > u$
- For small $\mathcal{A}$, large values of $\bar{R}_{\mathcal{A}}$ given by convective events
- As $|\mathcal{A}|$ increases, non-convective events become more important
- To jointly model $\bar{R}_{\mathcal{A}}$ for small and large $\mathcal{A}$, need a model for $Y(s)$ that can capture both types of events

Mixture modelling - Motivation

- Extreme precipitation driven by mixture of spatially localised, high intensity (convective) events and large-scale, low-intensity events (non-convective/frontal)
- In original work, we only model convective events - Recall $Y(s_O) > u$
- For small $\mathcal{A}$, large values of $\bar{R}_{\mathcal{A}}$ given by convective events
- As $|\mathcal{A}|$ increases, non-convective events become more important
- To jointly model $\bar{R}_{\mathcal{A}}$ for small and large $\mathcal{A}$, need a model for $Y(s)$ that can capture both types of events

Mixture modelling - Overview

- Assume Y is a mixture of "convective" and "non-covnective" processes:

$$\{Y_t(s) : s \in \mathcal{S}\} = \begin{cases} \{Y_{\mathcal{C},t}(s) : s \in \mathcal{S}\}, & \text{with prob. } p_{\mathcal{C}}(t) \in [0,1], \\ \{Y_{\mathcal{N},t}(s) : s \in \mathcal{S}\}, & \text{with prob. } 1 - p_{\mathcal{C}}(t), \end{cases}$$

- Current work assumes $p_{\mathcal{C}}(t)$ is constant for all t
- Identification algorithm developed at Hadley Centre, UK Met Office
- Separate marginal and extremal dependence model for each process

Mixture modelling - Overview

- Assume Y is a mixture of “convective" and “non-covnective" processes:

$$\{Y_t(s) : s \in \mathcal{S}\} = \begin{cases} \{Y_{\mathcal{C},t}(s) : s \in \mathcal{S}\}, & \text{with prob. } p_{\mathcal{C}}(t) \in [0,1], \\ \{Y_{\mathcal{N},t}(s) : s \in \mathcal{S}\}, & \text{with prob. } 1 - p_{\mathcal{C}}(t), \end{cases}$$

- Current work assumes $p_{\mathcal{C}}(t)$ is constant for all t
- Identification algorithm developed at Hadley Centre, UK Met Office
- Separate marginal and extremal dependence model for each process

Mixture modelling - Overview

- Assume Y is a mixture of “convective" and “non-covnective" processes:

$$\{Y_t(s) : s \in \mathcal{S}\} = \begin{cases} \{Y_{\mathcal{C},t}(s) : s \in \mathcal{S}\}, & \text{with prob. } p_{\mathcal{C}}(t) \in [0,1], \\ \{Y_{\mathcal{N},t}(s) : s \in \mathcal{S}\}, & \text{with prob. } 1 - p_{\mathcal{C}}(t), \end{cases}$$

- Current work assumes $p_{\mathcal{C}}(t)$ is constant for all t
- Identification algorithm developed at Hadley Centre, UK Met Office
- Separate marginal and extremal dependence model for each process

Mixture modelling - Overview

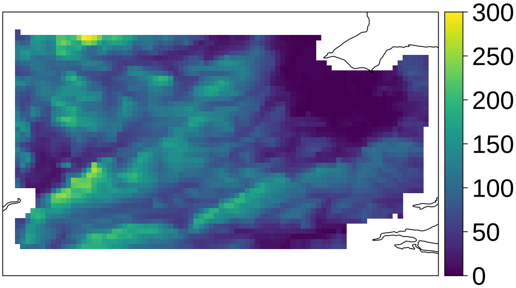
- Assume Y is a mixture of "convective" and "non-covnective" processes:

$$\{Y_t(s) : s \in \mathcal{S}\} = \begin{cases} \{Y_{\mathcal{C},t}(s) : s \in \mathcal{S}\}, & \text{with prob. } p_{\mathcal{C}}(t) \in [0,1], \\ \{Y_{\mathcal{N},t}(s) : s \in \mathcal{S}\}, & \text{with prob. } 1 - p_{\mathcal{C}}(t), \end{cases}$$

- Current work assumes $p_{\mathcal{C}}(t)$ is constant for all t
- Identification algorithm developed at Hadley Centre, UK Met Office
- Separate marginal and extremal dependence model for each process

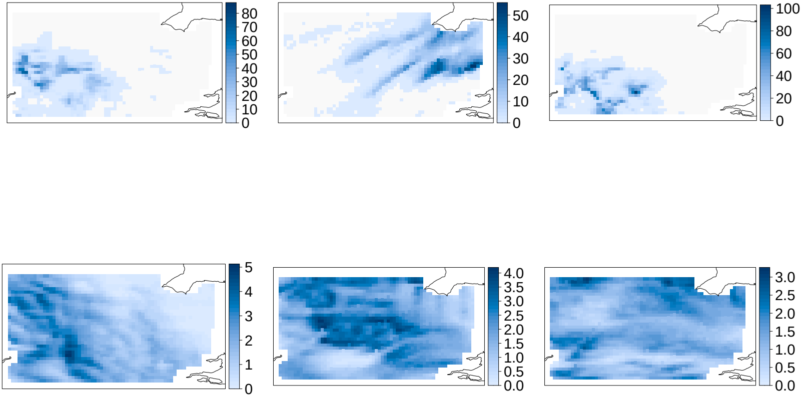
Data

- Same model as previously, but higher spatial resolution $(2.2km)^2$ and dimension $d = 7526$
- Larger spatial domain

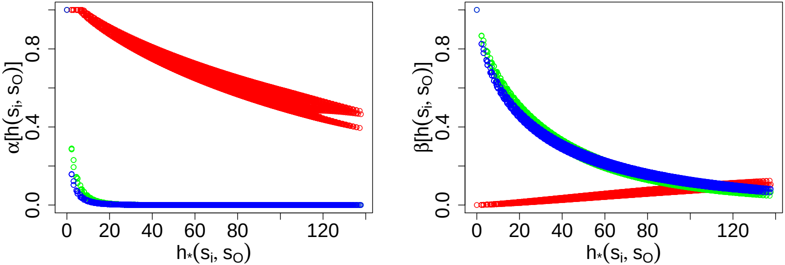


Example observations

Top: Convective. Bottom: Non-convective



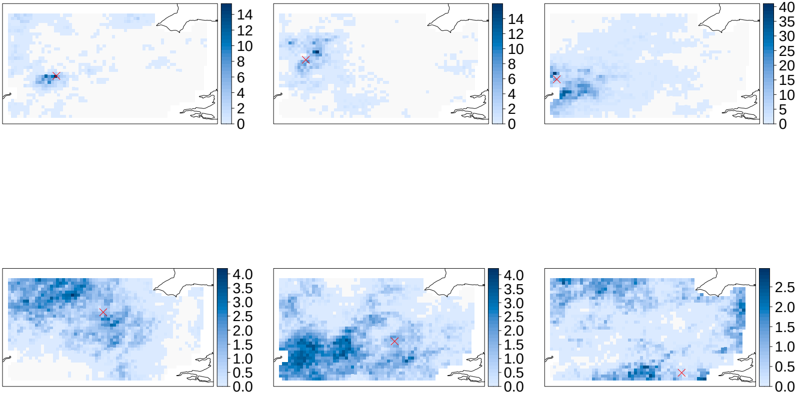
Comparison of fitted dependence models



- Green: Convective. Red: Non-convective.
- Observe much faster decay of extremal dependence for convective events.
- $\alpha(h) = 1$ for $\mathcal{N}$ and $h \leq 6.5km$

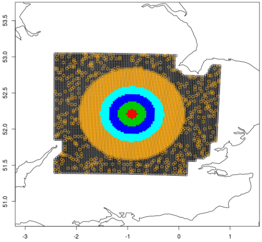
Simulated fields

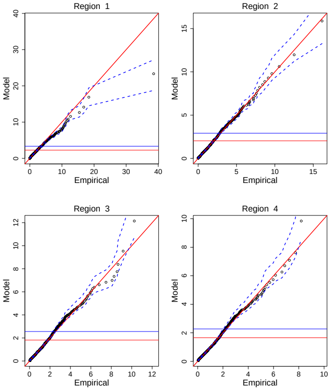
Top: Convective. Bottom: Non-convective



Aggregate diagnostics

- Regions 1 - 4, areas increasing from 179km^2 to 6200km^2
- The proportion of frontal events that contribute to $\bar{R}_{\mathcal{A}}$ above its 99.5% quantile increases from 0.4% to 4%





Extensions

- (Current) Incorporating temporal non-stationarity into extremal dependence functions - Quantify the effect of global warming on $\bar{R}_{\mathcal{A}}$ and two precipitation "types"
- Temporal extremal dependence
- Probabilistic clustering of events/more mixture components
- Spatial non-stationarity in α, β and $Z(s|s_O)$

References

Heffernan, J. E. and Tawn, J. A. (2004).
A conditional approach for multivariate extreme values (with discussion).
Journal of the Royal Statistical Society: Series B (Statistical Methodology), 66(3):497–546.

Richards, J., Tawn, J. A., and Brown, S. (2021).
Modelling extremes of spatial aggregates of precipitation using conditional methods.
arXiv preprint arXiv:2102.10906.

Wadsworth, J. L. and Tawn, J. A. (2019).
Higher-dimensional spatial extremes via single-site conditioning.
arXiv e-prints, arXiv:1912.06560.

Wood, S. (2006).
Generalized Additive Models: An Introduction with R.
Chapman & Hall/CRC Texts in Statistical Science. Taylor & Francis.

Youngman, B. D. (2019).
Generalized additive models for exceedances of high thresholds with an application to return level estimation for US wind gusts.
Journal of the American Statistical Association, 114(528):1865–1879.

Thanks for your attention!