Extreme Value Theory Notes

Patrick O'Toole

August 26, 2024

Contents

1 Introduction

Limit number of warnings from chktex

Replace notes in text with calls to mynote

Add introduction, including stuff from Coles book?

number equations

Remove numbers for unreferenced equations

Tidy bibtex references

Include authors names rather than numbering for references

2 Univariate extremes

3 Conditional extremes model

Look back into page 500 and equation 1.5, needed here?

Description of model comes from Heffernan and Tawn [2004], with the addition of using Laplace margins and constraining the normalising function parameters coming from Keef et al. [2013a].

3.1 Data

Uses beta for GPD scale, but beta used for normalising function, and want sigma for moments of Z, need to have consistent notation

- Continuous vector variable $\boldsymbol{X} = (X_1, X_2, \dots, X_d)$.
- Want to estimate $\mathbb{P}(\boldsymbol{X} \in \boldsymbol{C})$, where $\boldsymbol{C}$ is an extreme set such that $\forall \boldsymbol{X} \in \boldsymbol{C}$, at least one component of $\boldsymbol{X}$ i extreme.
- C_i corresponds to the part of C for which X_i is the largest component of $\boldsymbol{X}$, by quantiles of the marginal distribution.
- $C_i = C \cap \left\{\boldsymbol{x} \in \mathbb{R}^d : F_{X_i}(x_i) > F_{X_j}(x_j); j = 1, \dots, d; j \neq i\right\}$ for $i = 1, \dots, d.$, where F_{X_i} is the marginal distribution function of X_i.
- Ignore subsets $C \cap \left\{\boldsymbol{X} \in \mathbb{R}^d : F_{X_i}(X_i) = F_{X_j}(x_j) \text{for some} j \neq i\right\}$, as they are null sets.
- C is an extreme set if all x_i-values in non-empty C_i fall in upper tail of F_{X_i}, i.e. if $\nu_{X_i} = \inf_{\boldsymbol{x} \in C_i}(x_i)$, then $F_{X_i}(\nu_{X_i})$ is close 1 for $i = 1, \dots, d$, so

Improve equation referencing etc

$$\mathbb{P}(\boldsymbol{X} \in \boldsymbol{C}) = \sum_{i=1}^{d} \mathbb{P}(\boldsymbol{X} \in C_i) = \sum_{i=1}^{d} \mathbb{P}(\boldsymbol{X} \in C_i \mid X_i > \nu_{X_i})\mathbb{P}(X_i > \nu_{X_i})$$

- red probability is estimated with marginal extreme value model, while the blue probability is estimated using an extreme value model for the dependence structure.

3.2 Marginal extremes model

- Model marginal tail of X_i with Generalised Pareto Distribution (GPD): where u_{X_i} is a high threshold, ξ_i is the shape parameter, σ_i is the scale parameter, and $x_+ = \max(x, 0)$.
- Require a model for complete marginal distribution F_{X_i} of X_i, so need to describe all X_j values that can occur with any large X_i value, which leads to the following piecewise semiparametric model:
$$\hat{F}_{X_i}(x) = \begin{cases} 1 - \{1 - \tilde{F}_{X_i}(u_{X_i})\}\left\{1 + \xi_i(x - u_{X_i})/\sigma_i\right\}_+^{-1/\xi_i} & \text{if } x > u_{X_i} \\ \tilde{F}_{X_i}(x) & \text{if } x \leq u_{X_i} \end{cases}$$
where $\tilde{F}_{X_i}$ is the empirical distribution function of the X_i values.
- This gives us estimates of $\mathbb{P}(X_i < \nu_{X_i})$.

3.3 Marginal transformation

Describe need for marginal transformation to estimate dependence model

3.3.1 Gumbel transformation

$$\begin{aligned} Y_i &= -\log[-\log\{\hat{F}_{X_i}(X_i)\}], i = 1, \dots, d \\ &= t_i(X_i; \phi_i, \tilde{F}_{X_i}(X_i)) \\ &= t_i(X_i), \end{aligned}$$

where $\phi_i = (\sigma_i, \xi_i)$ are marginal parameters.
This gives $\mathbb{P}(Y_i \leq y) = \exp(-\exp(-y)) \implies \mathbb{P}(Y_i > y) \sim \exp(-y)$ as $y \to \infty$, so Y_i has an exponential upper tail.

3.3.2 Laplace transformation

The Laplace transformation detailed in Keef et al. [2013a] is given by

$$Y_i = \begin{cases} \log\{2F_{X_i}(x_i)\} & \text{for } X_i < F_{X_i}^{-1}(0.5) \\ -\log\{2[1 - F_{X_i}(x_i)]\} & \text{for } X_i \geq F_{X_i}^{-1}(0.5) \end{cases}$$

which means that

$$\mathbb{P}(Y_i \leq y) = \begin{cases} \exp(y)/2 & \text{for } y < 0 \\ 1 - \exp(-y)/2 & \text{for } y \geq 0 \end{cases}$$

so that both tails of Y_i are exponential, and so for any $u > 0$, the distribution of $Y_i - u \mid Y_i > u$ and $(-Y_i + u) \mid Y_i \leq -u$ are exponential with mean 1. This greatly simplifies the normalising functions seen in section 3.4.2, as for Gumbel margins a more complex normalising function is required for negatively associated variables.

3.4 Asymptotic dependence

- $$\lim_{y \to \infty} \{\mathbb{P}(\boldsymbol{Y}_{-i} \mid Y_i > y)\} = \begin{cases} 0 & \text{for asymptotic independence} \\ \neq 0 & \text{for asymptotic dependence,} \end{cases}$$

  where $\boldsymbol{Y}_{-i} = (Y_1, Y_2, \dots, Y_{i-1}, Y_{i+1}, \dots, Y_d)$.
- Existing methods for multivariate extremes (e.g. max-stable processes, copulas) can only model $\mathbb{P}(\boldsymbol{X} \in C)$ under asymptotic dependence.

revisit this and talk a bit more about it!

3.4.1 Limit assumption

- For each Y_i, want to estimate $\mathbb{P}(Y_{-i} \leq y_{-i} \mid Y_i = y_i)$ as $y \to \infty$.
- We require the limiting distribution to be non-degenerate for all margins (see section2)
- Therefore, assume for every i that there are vector normalising functions $\boldsymbol{a}_{|i}(y_i), \boldsymbol{b}_{|i}(y_i), \in \mathbb{R} \to \mathbb{R}^{(d-1)}$ such that for fixed $\boldsymbol{z}_{|i}$,

  $$\lim_{y_i \to \infty} \{\mathbb{P}(\boldsymbol{Y}_{-i} \leq \boldsymbol{a}_{|i}(y_i) + \boldsymbol{b}_{|i}(y_i)\boldsymbol{z}_{|i} \mid Y_i = y_i)\} = \boldsymbol{G}_{|i}(\boldsymbol{z}_{|i})$$

  where all margins of $\boldsymbol{G}_{|i}$ are non-degenerate, so

  revisit, probably wrong

  $$\lim_{z \to \infty} \boldsymbol{G}_{j|i}(\boldsymbol{z}) = 1, \forall j \neq i$$

  (no mass at $+\infty$, some allowed at $-\infty$).
- Alternatively, the standardised variables

  $$\boldsymbol{Z}_{|i} = \frac{\boldsymbol{Y}_{-i} - \boldsymbol{a}_{|i}(y_i)}{\boldsymbol{b}_{|i}(y_i)}$$

  have the property that

  $$\lim_{y_i \to \infty} \{\mathbb{P}(\boldsymbol{Z}_{|i} \leq \boldsymbol{z}_{|i} \mid Y_i = y_i)\} = \boldsymbol{G}_{|i}(\boldsymbol{z}_{|i})$$
- Conditional on $Y_i > u_i$ as $u_i \to \infty$, $Y_i - u_i$ and $\boldsymbol{Z}_{|i}$ are independent in the limit with limiting marginal distributions being exponential and $G_{|i}$ respectively:

  $$\mathbb{P}(\boldsymbol{Z}_{|i} \leq \boldsymbol{z}_{|i}, Y_i - u_i = y_i \mid Y_i > u_i) \to G_{|i}(\boldsymbol{z}_i)\exp(-y) \text{ ,as } u_i \to \infty$$

- For each $j \neq i$,
$$Z_{j|i} = \frac{Y_j - a_{j|i}(y_i)}{b_{j|i}(y_i)} \sim G_{j|i}(z_{j|i}) \text{ given } Y_i = y \text{ as } y_i \to \infty$$
$\implies G_{j|i}$ is the marginal distribution of $G_{|i}$ associated with Y_j.

3.4.2 Normalisation

- Under Gumbel margins, normalising functions have simple form for class of positively associated variable, but more complex for negatively associated variables: where $\boldsymbol{\alpha}_{|i}, \boldsymbol{\beta}_{|i}, \boldsymbol{\gamma}_{|i}, \boldsymbol{\delta}_{|i}$ are vector constants, $\boldsymbol{\alpha}_{|i} \in [0,1], \boldsymbol{\beta}_{|i} \in (-\infty, 1], \boldsymbol{\gamma}_{|i} \in (-\infty, \infty), \boldsymbol{\delta}_{|i} \in [0,1]$.

Fix, missing curly bracket

- Under Laplace marginals, as in Keef et al. [2013a], the indicator term disappears, and we have that $\boldsymbol{\alpha}_{|i} \in [-1,1]$.
- $\alpha_{j|i}$ controls the level of association between Y_j and large Y_i, with positive and negative values indicating positive and negative asymptotic dependence, respectively, and values closer to 0 indicating stronger asymptotic independence.
- $\beta_{j|i}$ controls the spread ...

Fill in more formally

Add interpretation of alpha and beta from Keef paper (page 400)

3.5 Conditional dependence model

- $\boldsymbol{G}_{|i}$ is modelled nonparameterically as the empirical distribution of
$$\boldsymbol{Z}_{|i} = \frac{\boldsymbol{Y}_{-i} - \boldsymbol{\alpha}_{|i}(Y_i)}{(Y_i)^{\boldsymbol{\beta}_{|i}}}$$
- All d different conditional distributions are estimated separately.
- Our dependence model therefore is a multivariate semiparamettiv regression model of the form
$$\begin{aligned} \boldsymbol{Y}_{-i} &= \boldsymbol{a}_{|i}(y_i) + \boldsymbol{b}_{|i}(y_i)\boldsymbol{Z}_{|i} && (3.1) \\ &= \boldsymbol{\alpha}_{|i}(y_i) + \boldsymbol{Z}_{|i}^{\boldsymbol{\beta}_{|i}} && (3.2) \end{aligned}$$
with the use of Laplace margins and the definition of $a_{|i}$ and $b_{|i}$ from Keef et al. [2013a].

This is the Conditional Extremes dependence model, the most important thing to know about it!

on page 508, do I need to talk about Z hat? How G hat is its empirical distribution?

3.6 Extrapolation

- Can simulate from $\boldsymbol{X}|\boldsymbol{X}_i > \nu_{X_i}$ by simulating from $\boldsymbol{Y}|\boldsymbol{Y}_i > y_i$ and transforming back to $\boldsymbol{X}$ space, using the following algorithm:
  1. Simulate Y_i from the transformed marginal distribution conditional on exceeding $t_i(\nu_{X_i})$.
  2. Sample $\boldsymbol{Z}_{|i}$ from $\hat{G}_{|i}$, independently of Y_i.
  3. Obtain $\boldsymbol{Y}_{-i} = \boldsymbol{\alpha}_{|i}(Y_i) + (Y_i)^{\boldsymbol{\beta}_{|i}}\boldsymbol{Z}_{|i}$.
  4. Transform $\boldsymbol{Y} = (\boldsymbol{Y}_{-i}, Y_i)$ back to the original scale by using the inverse of the marginal transformation.
  5. the resulting vector $\boldsymbol{X}$ is a simulated value from $\boldsymbol{X} \mid X_i > \nu_{X_i}$.
- This algorithm can be used to estimate $\mathbb{P}(\boldsymbol{X} \in C_i \mid X_i > \nu_{X_i})$ by evaluating it as the long run proportion of the generated sample that falls in C_i.

3.7 Diagnostics

- Marginals are diagnosed as in univariate EVT (threshold with mean excess plot etc. , model fit assessed with probability and quantile plots)
- Dependence:
  - Normalised variable $\boldsymbol{Z}_{|i}$ must have stable distribution over range of Y_i used for estimation and evaluation.
  - Independence of $\boldsymbol{Z}_{|i}$ and Y_i given $Y_i > u_i$ for high threshold u_i is explored (similar to diagnostics for 'ordinary' linear model)
  - Can also use standard statistical tests for independence.

3.8 Inference

- Inference for marginal parameters $\boldsymbol{\psi}$ and dependence parameters θ is done stepwise (loss of efficiency deemed small, estimation methods much easier)

3.8.1 Marginal Estimation

- d univariate distributions estimated jointly assuming independence between components in LL:

$$\log\{L(\boldsymbol{\psi})\} = \sum_{i=1}^{d} \sum_{j=1}^{n_{u_{X_i}}} \log\{\hat{f}_{X_i}(x_{i|i,k})\}$$

  f_{X_i} is the density associated with semiparametric marginal $\hat{F}_{X_i}$, and $n_{u_{X_i}}$ is the number of exceedances of u_{X_i}.

- Equivalent to fitting GPD for each margin to excesses over marginal threshold
- Can estimate GPD using more complicated methods, such as extreme value GAMs and spatiotemporal models which allow for covariates and information borrowing between marginals.

3.8.2 Single conditional

- Want to estimate $\boldsymbol{\theta}_{|i}$ under minimal assumptions about $\boldsymbol{G}_{|i}$.
- Assume $Z_{|i}$ has two finite marginal moments, $\boldsymbol{\mu}_{|i}$ and $\boldsymbol{\sigma}_{|i}$ (Note: $\boldsymbol{\sigma}_{|i}$ is the vector of standard deviations of $\boldsymbol{Z}_{|i}$, separate to parameter of GPD)
- Therefore $\boldsymbol{Y}_{-i} \mid Y_i = y$ for $y > u_{Y_i}$ has vector mean and standard deviation

$$\begin{aligned}\boldsymbol{\mu}_{|i}(y) &= \boldsymbol{a}_{|i}(y) + \boldsymbol{\mu}_{|i}\boldsymbol{b}_{|i}(y),\\ \boldsymbol{\sigma}_{|i}(y) &= \boldsymbol{\sigma}_{|i}\boldsymbol{b}_{|i}(y),\end{aligned}$$

- (Note: $\boldsymbol{\mu}_{|i}(y)$ is the vector mean, while $\boldsymbol{\mu}_{|i}$ are it's respective parameters)
- $\implies$ $(\boldsymbol{\theta}_{|i}, \boldsymbol{\lambda}_{|i} = (\boldsymbol{\mu}_{|i}, \boldsymbol{\sigma}_{|i}))$ are parameters of multivariate regression model to be estimated.
- Assume components of $\boldsymbol{Z}_{|i}$ are independent and Gaussian, for simplicity and convenience.
- Independence assumption is reasonable as $\boldsymbol{\theta}_{|i}$ determines only the marginal behaviour of the conditional distribution.

Do I explicitely define theta and psi as parameters of G and marginals anywhere?

Will have to change one of these, see Christian's paper (he uses psi and nu)

- The objective function for the point estimation of $(\boldsymbol{\theta}_{|i}, \boldsymbol{\lambda}_{|i})$ is

$$Q_{|i}(\boldsymbol{\theta}_{|i}, \boldsymbol{\lambda}_{|i}) = \sum_{j \neq i} \sum_{k=1}^{n_{u_{Y_i}}} \left[\log\{\sigma_{j|i}(y_{i|i,k})\} + \frac{1}{2} \left\{ \frac{y_{j|i,k} - \mu_{j|i}(y_{j|i,k})}{\sigma_{j|i}(y_{i|i,k})} \right\}^2 \right]$$

- Maximise jointly w.r.t $\boldsymbol{\theta}_{|i}$, $\boldsymbol{\lambda}_{|i}$, to obtain $\hat{\boldsymbol{\theta}}_{|i}$ with $\hat{\boldsymbol{\lambda}}_{|i}$ being nuisance parameters.
- For Gumbel margins, must estimate dependence model in two steps, first fixing $\gamma_{j|i} = \delta_{j|i} = 0$, then estimating both if the indicator function in equation hmm is satisfied.

reference correct equation

3.8.3 All conditionals

- Falsely assume independence between different conditional distributions:

$$\hat{Q}(\boldsymbol{\theta}, \boldsymbol{\lambda}) = \sum_{i=1}^{d} Q_{|i}(\boldsymbol{\theta}_{|i}, \boldsymbol{\lambda}_{|i})$$

- False assumption above approximates the pseudo-likelihood as marginal density of $\hat{Y}_{-i}$ and conditional density of $\hat{Y}_{-i} \mid Y_i = y_i$ when $y_i < u_{Y_i}$ negligibly influence the shape of the pseudo-likelihood.
- If variables all mutually asymptotically independent, then for sufficiently large u_{Y_i} each datum will exceed at most one threshold, so independence assumption will be satisfied.

3.9 Uncertainty estimation (bootstrap)

- Uncertainty comes from semiparametric marginal models, normalising functions and distributions of residuals.
- Semiparametric bootstrap used to evaluate standard errors of model parameter estimates and other estimated parameters such as $\mathbb{P}(\boldsymbol{X} \in \boldsymbol{C})$.
- Assume marginal and dependence thresholds are fixed, therefore uncertainty due to threshold selection not accounted for.
- Three stages: data generation under fitted model, estimation of model parameters, and derivation of estimate of parameters linked to extrapolation.
- Two-step sampling algorithm used for data generation to replicate both the marginal and dependence features of the data.
- ...

Finish this section

3.10 Return level

When multivariate set $\boldsymbol{C}$ is described by single parameter nu (i.e. $C = C(\nu)$), the return level ν_p for event with probability p is defined as

$$\mathbb{P}(\boldsymbol{Y} \in C(\nu_p)) = p$$

3.11 New constraints

Some new constraints for the conditional extremes model were introduced in Keef et al. [2013a].

- The coefficient of tail dependence between the pair of variables (X_i, X_j) is given by

$$\chi_{ij}(p, q) = \mathbb{P}\{X_j > F_j^{-1}(q) \mid X_i > F_i^{-1}(p)\} \text{ for } p, q \in (0, 1)$$

- The limiting positive dependence between two variables is given by

$$\chi_{ij}^{+} \lim_{p \to 1} \chi_{ij}(p, p)$$

, while the limiting negative dependence is

$$\chi_{ij}^{-} = \lim_{p \to 1} 1 - \chi_{ij}(1-p, p)$$

- Asymptotic positive dependence between pair (X_i, X_j) gives $\chi_{ij}^{+} > 01$, while asymptotic independence gives $\chi_{ij}^{+} = 0$.
- Similarly, asymptotic negative dependence gives $\chi_{ij}^{-} > 0$, while asymptotic independence gives $\chi_{ij}^{-} = 0$.
- We must preserve the bounds implied by the asymptotic dependence through a stochastic ordering of the conditional distributions of $Y_j \mid Y_i = y$ for large y associated with asymptotic negative dependence, asymptotic independence, and asymptotic positive dependence, respectively.
- Otherwise, the resulting joint probabilities can exceed the marginal probabilities, i.e.

$$\hat{\mathbb{P}}(X_i > F_i^{-1}(p), X_j > F_j^{-1}(q)) > \max(1-p, 1-q)$$

- Let the q^{th} conditional quantile of $Y_j \mid Y_i = y$ for large y be $y_{j|i}(q)$ and quantiles under the assumption of asymptotic positive and negative dependence be $y_{j|i}^{+}(q)$ and $y_{j|i}^{-}(q)$, respectively. Then

$$y_{j|i}^{-}(q) \leq y_{j|i}(q) \leq y_{j|i}^{+}(q) \tag{3.3}$$

for

$$\begin{aligned} y_{j|i}^{-}(q) &= -y + \boldsymbol{Z}_{j|i}^{-}(q), \\ y_{j|i}(q) &= \alpha_{j|i} y + y^{\beta_{j|i}}(q), \\ y_{j|i}^{+}(q) &= y + \boldsymbol{Z}_{j|i}^{+}(q), \end{aligned}$$

where $\hat{G}_{j|i}^{-}\{Z_{j|i}^{-}(q)\} = \hat{G}_{j|i}\{Z_{j|i}(q)\} = \hat{G}_{j|i}^{+}\{Z_{j|i}^{+}(q)\} = q$, where the $\hat{G}_{j|i}^{-}, \ldots$ are the estimated empirical distributions of $\boldsymbol{Z}_i$ for $Y_i > u$ under the assumption of asymptotic negative dependence, etc.

- Under the assumption of the asymptotic dependence for (Y_i, Y_j) with $Y_i > u$, $Z_{j|i}^{+}(q)$ is the empirical q^{th} quantile of $Z_{j|i}^{+} = Y_j - Y_i$ for $Y_i > u$.
- Similarly, $\ldots Z_{j|i}^{-}(q) = Y_j + Y_i$.
- Under the conditional extremes model, we estimate $Z_{j|i}(q)$ as the empirical q^{th} quantile of $Z_{j|i} = (Y_j - \alpha_{j|i} Y_i)/Y_i^{\beta_{j|i}}$
- The theorem governing this ordering constraint is given by:
For $\nu > u$, the ordering constraint 3.3 holds for all $y > \nu$ iff both Case 1 and II hold. Fix theorem environment
Case 1: Either

$$\alpha_{j|i} \leq \min\{1, 1 - \beta_{j|i} Z_{j|i}(q) \nu^{\beta_{j|i}-1}, 1 - \beta_{j|i} Z_{j|i}(q) + \nu^{-1} Z_{j|i}^{+}(q)\}$$

or Finish theorem! Actually painful to type out lol

- We only impose constraints on extrapolation to give greatest flexibility $\implies$ take v to be above the maximum observed Y_i.
- The constraints are built into the inference by having the profile likelihood for $(\alpha_{j|i}, \beta_{j|i})$ obtained by having the likelihood equal 0 if Theorem 1 is not satisfied.
- Also reduces variance of estimators by removing inconsistent estimates.

3.11.1 Application

Do this section

3.12 Comparisons to other methods

Do this section, try summarise, no need for too much detail can also use what other papers citing this one have said!

In Tawn et al. [2018], the conditional extremes model is compared with other commonly used multivariate extremes models, namely max-stable, Pareto and Gaussian processes ...

3.13 Spatial conditional extremes

Write this section!

In Wadsworth and Tawn [2018], the conditional extremes model is extended to the spatial domain.

Also talk about spatio-temporal conditional extremes from Simpson INLA paper?

4 Applications of conditional extremes model

There have been many papers which have used the Conditional Extremes model in the context of multivariate extremes.

Estimating the probability of widespread flood events In Keef et al. [2013b], ...

Joint modelling of extreme ocean environments incorporating covariate effects

Modelling the effect of the El Niño-Southern Oscillation on extreme spatial temperature events over Australia

High-dimensional modeling of spatial and spatio-temporal conditional extremes using INLA and Gaussian Markov random fields

Model-based inference of conditional extreme value distributions with hydrological applications

5 Bayesian spatial clustering for extremes

Make this a subsection of a clustering section!

This algorithm for Bayesian spatial clustering of (hydrological) extremes is detailed in Rohrbeck and Tawn [2021].

- Clustering is done for two main reasons:
  - interpretation
  - improve inference by pooling information over similarly distributed variables.
- The first is usually done for extremes by fitting multiple marginal GPDs and applying generic clustering techniques (k-means, k-mediods, etc) on some summary statistic (such as the scale parameter, σ).
- The second often uses hierarchical modelling. Naturally in extreme analyses we have a lack of data, so pooling similar sites etc. for parameter estimation is highly advantageous.
- Existing spatial clustering focuses on either the marginal distributions or the dependence structure.
- This Bayesian clustering algorithm combines both ((1) GPD and (2) χ) into likelihood, with reversible jump MCMC algorithm used to estimate cluster allocation and cluster specific marginal parameters.

5.1 Model

5.1.1 Data

- K sites with spatial locations $s_1, \ldots, s_k \in \mathbb{R}^2$.
- For areal data, s_i is the centroid of the k^{th} areal unit.
- For geostatistical data, s_i is the point location of k^{th} site.
- Distance $d_{k,k'} \geq 0$ between sites k and k'.
- Declustering performed for each site, as data is seasonal and spatio-temporally dependent - only use highest observation per subperiod for which sitewise data is broken into..
- $R_{k,1}, \ldots, R_{k,t}$ denotes the time series for site k after declustering.
- Assumed independent $\forall t \neq t'$.
- K latent variables $\boldsymbol{Z} = (Z_1, \ldots, Z_k)$, where Z_k is the latent variable representing the cluster membership for site k.
- Let $J \in \{1, \ldots, K\}$ represent the number of clusters (so J is also a random variable!).
- Cluster based on (1) similar marginal distributions and (2) spatial dependence (represented with $\chi_{k,k'}$) being greater between sites in the same cluster than sites in different clusters.

5.1.2 Marginal Model

- $R_{k,t} - u_k \mid R_{k,t} > u_k \sim \text{GPD}(\psi_k, \nu_k)$, so we have site-specific marginal parameters and thresholds.
- In clustering, want cluster-specific, rather than site-specific GPD parameters, i.e.

$$R_{k,t} - u_k \mid (Z_k = j, R_{k,t} > u_k) \sim \text{GPD}(\sigma_j, \xi_j).$$

Note here that we still have site-specific thresholds.

Need to unify notation here

- Parameters of marginal model given $\boldsymbol{Z}$ denoted by $\boldsymbol{\theta}_m^{(J)} = \{\boldsymbol{\sigma}^{(J)}, \boldsymbol{\psi}^{(J)}\}$, where $\boldsymbol{\sigma}^{(J)} = (\sigma_1, \ldots, \sigma_J)$ and $\boldsymbol{\psi}^{(J)} = (\psi_1, \ldots, \psi_J)$.

have J superscript within bm call

5.1.3 Dependence Model

Here, we will replace the use of χ with the semiparametric conditional extremes model for dependence in equation 3.1

- Rather than model full joint distribution over extreme events at sites, model $\chi_{k,k'}$ for pairwise extremal dependence ($\chi_{k,k'}$ defined above)

Have section on multivariate extremes which has a definition for chi

- $\chi_{k,k'} \in [0,1]$ gives the limit probability of site k observing an extreme event given site k' recording one.
- Constrain $\chi_{k,k'}$, conditional on $\boldsymbol{Z}$ such that expected value is larger within clusters than for sites in different clusters:

$$\mathbb{E}(\chi_{k,k'} \mid Z_k = Z_{k'}) \geq \mathbb{E}(\chi_{k,k'} \mid Z_k \neq Z_{k'}) \tag{5.1}$$

- Further constrain so that $\chi_{k,k'}$ decreases with increasing distance between sites $d_{k,k'}$, with exponential decay which is also faster for sites in different clusters:

$$\mathbb{E}(\chi_{k,k'} \mid Z_k = j, Z_{k'} = j') = \begin{cases} \exp(-\gamma_j d_{k,k'}) & \text{if } Z_k = Z_{k'} = j \\ \exp(-\gamma_0 d_{k,k'}) & \text{if } Z_k \neq Z_{k'}, \end{cases} \tag{5.2}$$

where $\gamma_0 > \max(\gamma_1, \ldots, \gamma_J) \geq 0$, which is ensured by introducing parameters $(\epsilon_1, \ldots, \epsilon_J), \epsilon_j \geq 0$ and constraining the cluster specific decay factors s.t. $\log(\gamma_j) = \log(\gamma_0) - \epsilon_j$.

- $\chi_{k,k'} \mid \boldsymbol{Z} \in [0,1]$ may differ between two pairs of sites and within the same cluster with same $d_{k,k'}$ (due to factors like topology), so we choose a beta distribution model with

Ask Christian about use of Beta distribution here

$$\chi_{k,k'} \mid \boldsymbol{Z} \sim \begin{cases} \text{Beta}\left(\frac{\beta \exp(-\gamma_j d_{k,k'})}{1-\exp(-\gamma_j d_{k,k'})}, \beta\right) & \text{if } Z_k = Z_{k'} = j \\ \text{Beta}\left(\frac{\beta \exp(-\gamma_0 d_{k,k'})}{1-\exp(-\gamma_0 d_{k,k'})}, \beta\right) & \text{if } Z_k \neq Z_{k'} \end{cases} \tag{5.3}$$

which has expectation equal to 5.1.

- $\beta > 0$ is inversely proportional to the variance of $\chi_{k,k'}$.
- Therefore, dependence parameters are $\boldsymbol{\theta}_D^{(J)} = \gamma_0, \boldsymbol{\epsilon}^{(J)}, \beta\}$.

Why not gamma j? Do we only need to estimate the epsilons that relate them to gamma 0?

- $\chi_{k,k'}$ only considered for adjacent sites.
- For geostatistical data, we derive the Voronoi partition of area and define sites as being adjacent if Voronoi cells are.

Could be expanded with SPDE approach from INLA? Interesting compromise though

5.2 Inference

- Bayesian inference used to estimate $J, \boldsymbol{Z}, \boldsymbol{\theta}_m^{(J)}, \boldsymbol{\theta}_D^{(J)}$, using declustered data $\boldsymbol{D} = \{r_{k,1}, \ldots, r_{k,t}\}; k = 1, \ldots, K$.
- For marginal, data is marginal exceedances over threshold while for dependence the ranks of variables are used.
- Because of this difference in data, inference for both sets of parameters is largely independent, and we can approximate the joint likelihood using the independent decomposition:

$$L(\boldsymbol{\theta}_m^{(J)}, \boldsymbol{\theta}_D^{(J)} \mid \boldsymbol{D}, \boldsymbol{Z}) = L_m(\boldsymbol{\theta}_m^{(J)} \mid \boldsymbol{D}, \boldsymbol{Z}) L_D(\boldsymbol{\theta}_D^{(J)} \mid \boldsymbol{D}, \boldsymbol{Z}). \tag{5.4}$$

5.2.1 Marginal component

- Assuming thresholded data independent over all sites:

$$L_m(\boldsymbol{\theta}_m^{(J)} \mid \boldsymbol{D}, \boldsymbol{Z}) = \prod_{k=1}^{K} \prod_{\{t: r_{k,t} > u_k\}} \frac{1}{\sigma_{Z_k}} \left(1 + \xi_{Z_k} \frac{r_{k,t} - u_k}{\sigma_{Z_k}}\right)_+^{-1/\xi_{Z_k} - 1}$$

Ask Christian why the subscript is Zk here, rather than just k?

- However, assuming spatial independence is not a valid assumption, as severe weather events usually affect a number of sites simultaneously.

There could be an option to include this additional constraint to the marginal component of the likelihood in an R package

- General theory for asymptotic distribution of MLE $\hat{\boldsymbol{\theta}}_m^{(J)}$ for GPD parameters under independence assumption is

$$\sqrt{T}(\hat{\boldsymbol{\theta}}_m^{(J)} - \boldsymbol{\theta}_m^{(J)}) \sim \mathrm{N}(0, \Sigma = H(\boldsymbol{\theta}_m^{(J)})^{-1} V(\boldsymbol{\theta}_m^{(J)}) H(\boldsymbol{\theta}_m^{(J)})^{-1}),$$

where $\boldsymbol{\theta}_m^{(J)}$ are the true parameter, H denotes the Fischer information (see paper for full definition).

Using $L_m(\boldsymbol{\theta}_m^{(J)} \mid \boldsymbol{D}, \boldsymbol{Z})$ would underestimate the variance of $\boldsymbol{\theta}_m^{(J)}$ due to spatial dependence.

Adjust likelihood around it's mode:

$$L_m^{\text{adj}}(\boldsymbol{\theta}_m^{(J)} \mid \boldsymbol{D}, \boldsymbol{Z}) = L_m(\boldsymbol{\theta}_m^{(J)} + \boldsymbol{B}(\boldsymbol{\theta}_m^{(J)} - \hat{\boldsymbol{\theta}}_m^{(J)}) \mid \boldsymbol{D}, \boldsymbol{Z}),$$

where the $2J \times 2J$ matrix B depends on $\boldsymbol{Z}$ and is

Fix

B is a block matrix of $J 2 \times 2$ blocks for each clustering, allowing for efficient computation.

Take $L_m(\boldsymbol{\theta}_m^{(J)} \mid \boldsymbol{D}, \boldsymbol{Z}) = L_m^{\text{adj}}(\boldsymbol{\theta}_m^{(J)} \mid \boldsymbol{D}, \boldsymbol{Z})$.

i.e. use the adjusted likelihood hereafter for the marginal component of the likelihood

5.2.2 Dependence component

Assume $\exists \tilde{u} \in [0,1)$ s.t. $\chi_{k,k'}(u) = \chi_{k,k'} \forall \tilde{u} < u < 1$, i.e. it's equal to it's limit form, for all pairs of sites (k, k'). Let

$$mathcalQ_{k,k'} = \{t : \hat{F}_k(r_{k',t} > \tilde{u}\}$$

be the set of times when there is an exceedance of the quantile threshold at both sites k and k'.

The cardinality of this set is $Q_{k,k'} = |\mathcal{Q}_{k,k'}|$.

Let $P_{k,k'} = \{t \in \mathcal{Q}_{k,k'} : \hat{F}_k(r_{k,t}) > \tilde{u}\}$

Ask Christian how P is different to Q?

$\implies P_{k,k'} \sim \text{Binomial}(Q_{k,k'}, \chi_{k,k'})$.

Estimate for $\chi_{k,k'}$ is $\hat{\chi}_{k,k'} = P_{k,k'}/Q_{k,k'}$, which is the proportion of exceedances of the $100\tilde{u}\%$ for site k' that also exceed this quantile threshold for site k.

Combining this and Beta model for $\chi_{k,k'} \mid Z$ in equation 5.3, and then integrating over $\chi_{k,k'}$, we obtain that

$$P_{k,k'} \mid \boldsymbol{D}, \boldsymbol{Z} \sim \begin{cases} \text{Beta-binomial}\left(Q_{k,k'}, \frac{\beta}{\exp \gamma_j d_{k,k'}}, \beta\right) & \text{if } Z_k = Z_{k'} = j \\ \text{Beta-binomial}\left(Q_{k,k'}, \frac{\beta}{\exp \gamma_0 d_{k,k'}}, \beta\right) & \text{if } Z_k \neq Z_{k'} \end{cases} \tag{5.5}$$

This model to hold only for pairs of adjacent sites.

Denote the density function in 5.5 by g, then for a pair of adjacent sites (k, k') the likelihood contribution to $L_D(\boldsymbol{\theta}_D^{(J)} \mid \boldsymbol{D}, \boldsymbol{Z})$ is

$$L_D\left(\boldsymbol{\theta}_D^{(J)} \mid \boldsymbol{D}, \boldsymbol{Z}\right) = \left[g\left(P_{k,k'} \mid Q_{k,k'}, \boldsymbol{Z}, \boldsymbol{\theta}_D^{(J)}\right) \times g\left(P_{k',k} \mid Q_{k',k}, \boldsymbol{Z}, \boldsymbol{\theta}_D^{(J)}\right)\right]^{0.5},$$

as we have two estimates for $\chi_{k,k'}$ which contain almost exactly the same information.

Under the assumption of independence for distinct pairs,

$$L_D(\boldsymbol{\theta}_D^{(J)} \mid \boldsymbol{D}, \boldsymbol{Z}) = \prod_{k \sim k'} L_D^{k,k'}$$

Likelihood is misspecified, . . .

Complete

5.3 Priors

- Priors required for $J, \boldsymbol{Z} \mid J, \boldsymbol{\theta}_m^{(J)}, \boldsymbol{\theta}_D^{(J)}$.
- $J \geq 1 \implies J - 1 \sim \text{Poisson}(\kappa), \kappa \sim \text{Gamma}(1, 0.001)$ (weakly informative).
- Clusters must be contiguous over space only give positive mass in prior to contiguous clusters.
- $\boldsymbol{C}^{(J)} = \{C_1, \ldots, C_J\} \in \{1, \ldots, K\}, C_i \neq C_j$ if $i \neq j$ are centres of J clusters, with each k assigned to the closest cluster centre in terms of distance:

$$Z_k \mid \boldsymbol{C}^{(J)} = \underset{j \in \{1,\ldots,J\}}{\arg\min}\ d_{k,C_j} \tag{5.6}$$

- site k assigned to cluster with lowest index if multiple centres minimise distance to the site (not well-defined otherwise).
- Can assign prior to $\boldsymbol{Z} \mid J$ via $C^{(J)}$, which is given a uniform prior with

$$\mathbb{P}(C^{(J)} \mid J) = \frac{(K - J)!}{K!}$$

- With independent conjugate priors for the hyperparameters of both, we have that for $\boldsymbol{\theta}_m^{(J)}$:
  - $\sigma_j \sim \text{LogNormal}(\mu^\sigma, \theta^\sigma), \mu^\sigma \sim \text{Normal}(0, 1), \theta^\sigma \sim \text{Inverse-Gamma}(1, 0.1)$.
  - $\xi_j \sim N(\mu^\xi, \theta^\xi), \mu^\xi \sim \text{Normal}(0, 0.2), \theta^\xi \sim \text{Inverse-Gamma}(1, 0.1)$,

  Could use penalised complexity prior for ξ, which would penalise the likelihood of having values differing greatly from 0 (implemented in INLA)

  and for $\boldsymbol{\theta}_D^{(J)}$:
  - $\epsilon_j \sim \text{Exp}(\theta^\epsilon), \gamma_0 \sim \text{Exp}(0.001), \beta \sim \text{Exp}(0.001), \theta^\epsilon \sim \text{Gamma}(5, 2)$.

5.4 Reversible jump MCMC Algorithm

Do this section!

- Want to sample from posterior distribution defined by the likelihood in 5.4 and the priors in 5.3.
- The dimension of the parameter space changes with J, so use a reversible jump MCMC algorithm.
- Given a current sample with J clusters, propose one of the following seven moves:

1. **Birth**: Introduce a new cluster centre C^* with parameters $\epsilon^*, sigma^*$ and ξ^*.
2. **Death** Remove one of the existing cluster centres $C_1, \ldots, {}_J$. Sites previously allocated to the removed cluster are assigned to the other $J-1$ clusters according to the distance criterion in equation 5.6.
3. **Shift** Move one cluster centre to nearby site which is not a cluster centre, update cluster labels $\boldsymbol{Z} \mid J$ according to equation 5.6.
4. **Sigma** Update $\boldsymbol{\sigma}^{(J)}$ for all clusters.
5. **Xi** Update $\boldsymbol{\xi}^{(J)}$ for all clusters.
6. **Chi** Update $\boldsymbol{\epsilon}^{(J)}, \gamma_0, \beta$.
7. **Hyper** Update hyperparameters $\kappa, (\mu^\sigma, \theta^\sigma), \left(\mu^\xi, \theta^\xi, \theta^\epsilon\right)$.

- For a birth move, a new cluster centre C^* is uniformly sampled from one of the $K-J$ sites which are not currently cluster centres.

  Cluster centres are therefore "mediods" in some sense, in that they are actually observed sites, rather than some aggregation between sites

- Index at which to insert C^* is also uniformly sampled.

Finish!

6 Other methods for extremal clustering

Similar to section 4, this section will detail other methods for clustering of extremes.

Clustering of maxima: Spatial dependencies among heavy rainfall in France In Bernard et al. [2013], a clustering algorithm for (weekly precipitation) maxima is proposed which uses (an empirical/nonparametric estimator for) the F-madogram, a type of variogram (estimates spatial correlation of spatial random field) for extremes, and PAM (partitioning around mediods).
Summary:

- bivariate vector $(M_i, M_j)^T$ assumed to follow a bivariate EVT distribution.
- A variogram of order p is defined as the moment of order p of the difference between maxima M_i and M_j, $E|M_i - M_j|^p$.
- The F-madogram d_{ij} is, amongst other things:
  1. an interpretable distance, thus forming a distance matrix over which to cluster,
  2. expressed in terms of the scalar "extremal coefficient" $V_{ij}(1,1)$, which gives information about the degree of dependence between M_i and M_j.
  3. forms a copula, and so is completely decoupled from marginal estimates, meaning there is no need to re-estimate a GEV distribution at each site, and it is not required to assume these maxima come from a GEV, but only that they lie in the domain of attraction of max-stable distribution.
- PAM preferred to k-means, for the following reasons:
  1. K-means averages over all cluster members. The average of normally distributed obs remain Gaussian, but this is not the case for maxima following an extremal distribution.
  2. Taking mediods ensures that the cluster centre is an actual observation, which allows the maxima to remain maxima and does not apply any averaging or smoothing.
- Interestingly, PAM not given geographical information, but still produces spatially coherent clusters.
- Choice of K (number of clusters) and assessment of clustering quality made through analysis of silhouette coefficient, which compares cluster tightness (small distance within cluster) with cluster dissociation/separation (clusters should be adequately distinct).

Partitioning into hazard subregions for regional peaks-over-threshold modeling of heavy precipitation Read paper!

Carreau et al. [2017] derives a model which uses covariates to estimate $\sigma(\boldsymbol{x})$ and conditional mixture of GPDs with subregions defined by constant shape parameters ξ_j, estimated through EM algorithm and partitioned into subregions, the number of which was determined via out-of-sample cross validation.
Summary:

- As $T \uparrow$, ξ becomes determining factor in return/hazard level estimates, particularly in how it determines the tail behaviour of the distribution.
- ξ estimation is difficult, as it is highly variable and can be influenced by many factors, such as the choice of threshold, the number of observations, etc.
- Approach here is to treat hazard level as piecewise constant, and partition region of interest into subregions over which to fit a conditional mixture of GPDs (can be seen as extension of Cooley et al. [2007] which uses two subregions).

- # regions sees bias-variance tradeoff; more regions means more ξ to estimate with less data (more variance), while less regions means greater bias as each subregion will represent the local characteristics of more locations (more groups often also harder to interpret).
- σ doesn't have same issue, so just estimated with $\sigma(\boldsymbol{x})$, i.e. a function of covariates.
- EM algorithm:
  - **E-step** estimates partition C, uses probability weighted moment estimators, U-statistics, kernel regression and k-means (complex, see paper)
  - **M-step** estimates $\mathbb{P}(C = j \mid \boldsymbol{x})$ probability of subregion membership used in mixture GPD, $\sigma(\boldsymbol{x}), \xi_j$ for each subregion j.
- Number of subregions chosen by out-of-sample CV using three different loss fuctions, including Anderson-Darling statistic (more sensitive to changes in the tails)

Clustering bivariate dependencies of compound precipitation and wind extremes over Great Britain and Ireland Vignotto et al. [2021] k-mediods algorithm of KL divergence between events characterised by risk function (sum/max) for Pareto transformed extreme wind and precipitation observations over Great Britian and Ireland.
Summary:

- **Data**: Weekly sum of precipitation and average of daily wind speed maxima used, as precipitation and wind speed extremes can be linked through storms with a lag of several days due to persistent weather patterns.
- **Mapping from bivariate to univariate space**: Word better
  - Marginal distributions (i.e. rain and wind speed at a single site) are transformed to standard Pareto distributions.
  - Risk function computed on Pareto scale $r : \mathbb{R}^2 \to \mathbb{R}$ used to define which points are extreme, where $r(x, y) = x + y$ or $r(x, y) = \max(x, y)$, mapping from bivariate to univariate space (taking (Pareto transformed) wind and precipitation estimates at each site and giving us a single estimate for each)

  is the Risk function only used to determine which points are extreme, with the original data bivariate data used to calculate KL divergence???

  Distance metric:
  - coefficient of tail dependence $\chi/\bar{\chi}$ cannot directly measure similarity of extremal behaviour of two bivariate random variables $\boldsymbol{X}^{(1)} = (X_1^{(1)}, X_2^{(1)})$ and $\boldsymbol{X}^{(2)} = (X_1^{(2)}, X_2^{(2)})$, i.e. rain and wind speeds at two different locations.
  - Instead, use the Kullback-Leibler divergence, which measures the difference between two probability distributions P and Q, and is defined as

    $$D_{KL}(P \mid Q) = \int p(x) \log\left(\frac{p(x)}{q(x)}\right) dx.$$

    KL divergence is shown to generalise the concept of χ, since ... Finish
  - Extreme points $\{R^{(j)} > q_u^{(j)}\}$ partitioned into $W = 3$ sets, one for co-occurring extremes and two for where data is extreme for only one variable (easily extended to multivariate case from bivariate).
  - Empirical proportions of data points belonging to each set used to estimate KL divergence between any two sites (i.e. how similar occurrence of extremes are at two sites), this giving distance/dissimilarity matrix over which to cluster.

- K-mediods algorithm clusters sites over KL divergences.
- Silhouette coefficient used to choose number of cluster and assess solution.
- Makes interesting conclusions about bivariate extremal behaviour of Ireland which will be useful for TFR report, but is different in it's use of gridded data, which underestimates extremal precipitation and wind speeds.

Modelling panels of extremes Dupuis et al. [2023] derives EM (really MM) algorithm for identifying group structure and group-specific model parameters for GEV distributed panel data. Summary:

- **Data**: data for individuals/locations i, time t (known as panel data, i.e. $X_{i\ t}$) which are GEV distributed.
- Because of the nature of extremes, both complete and no pooling results in poor parameter estimates, so some/"partial" pooling over similar/homogeneous locations desired.
- Use EM/EE algorithm and QML (where variance-covariance matrix may be misspecified) to iteratively estimate (i.e. estimation is disentangled for simplicity) group structure/assignment $\boldsymbol{\tau}$ and group-specific parameters $\boldsymbol{\theta}$ related to parameters of GEV through individual regression equations (so each GEV parameter estimated separately). The consistency of this algorithm is also proven.
- grouping is latent and "data-driven", rather than based on domain knowledge, and is done mainly for improved parameter estimation, as opposed to methods which derive distance matrices over which to perform e.g. k-mediods clustering.
- Stronger dependence in data helps group identification because it reduces the variance among individuals in the same group, but it gives worse quantile estimates (i.e. stronger dependence gives lower effective sample size $\implies$ less information $\implies$ greater variance in QML estimator).
- Simulation study performed, and applications to financial risk, extreme temperature and flood risk data shown to be effective (and better than when using groupings based on domain knowledge).

Similarity-based clustering for patterns of extreme values de Carvalho et al. [2023] uses k-means to cluster over bivariate cluster centroid of the extremal index and the heteroskedastic function, interpreted as the magnitude and frequency of extreme events, respectively. Summary (from reading course report):

- Heteroskedastic extremes violate assumption of IID observations, and may exhibit serial dependence or be drawn from different distributions.
- The **heteroskedastic function** c gives the frequency of extremes, with $c = 1$ defining "homoskedastic extremes". Defined as a limit which compares only the distribution tails, not imposing any assumption on centre of distributions.
- **Extremal index** γ **(the same as the shape parameter!)** is a scalar which controls the behaviour of a CDF in its right tail (i.e. it's rate of tail decay), and is often thought of as the inverse of the limiting mean cluster size).
- Nonparametric kernel-based estimator and Hill function provide respective estimates $\hat{c}, \hat{\gamma}$.
- Two quantities are jointly thresholded such that a specific quantile of observations is preserved.
- The level of bias towards one of these metrics is parameterised in this procedure, so that one can be favoured over the other in an analysis, as deemed necessary.

- Standard k-means clustering performed on bivariate cluster centroid $(\hat{c}, \hat{\gamma})$.

7 Bayesian extremes

Bayesian statistics has seen widespread adoption in the context of Extreme Value Theory. Naturally, as parameter and estimation uncertainty is high, the use of priors to incorporate expert knowledge, as well as hierarchical modelling to share information, is highly advantageous. Here, we will detail some of the methods used in Bayesian extremes.

Bayesian analysis of extreme values by mixture modeling Bottolo et al. [2003] defines exceedances over a given threshold as generated by a model characterised by a Poisson process $PP(\mu, \sigma, \xi)$, and derives a hierarchical mixture prior for each of these parameters which has (unknown, latent) parameter-specific group structuring (allowing for great flexibility) estimated through a Reversible Jump MCMC (RJMCMC) scheme similar to that of Rohrbeck and Tawn [2021].

- Data is exceedances over a threshold, which are modelled as a (Poisson) point process with parameters μ, σ, ξ.
- incorporating prior knowledge in context of extremes useful due to rarity of extremal data.
- Proposed Bayesian hierarchical model with (parameter-specific) grouping of type-effects, where grouping is latent and "data-driven", as in Dupuis et al. [2023].
- RJMCMC algorithm used to estimate group structure and group-specific parameters.
- Previous models assumed exchangeability of parameters meaning that types were treated symmetrically and correspondingly parameter estimates were shrunk towards common points.
- Under mixture priors, parameters are assumed to be i.i.d. according to some finite-mixture distribution with number of components w, corresponding weights k hyperparameters δ, and latent variable Z_i indicating the mixture component to which the parameter belongs.
- Hyperpriors perform partial-pooling between groups, with $k \sim \mathbb{P}(k^\eta = k)$ (can be uniform, fully specified or somewhere in between) for a PP parameter η and Dirichlet prior on $w^\eta \mid k^\eta$ (nice DAG for these in paper).

Bayesian spatial modeling of extreme precipitation return levels

- Abstract/summary:
  - Cooley et al. [2007] models **r-year return levels** (with uncertainty) for extreme precipitation in Colorado.
  - Separately hierarchical models for intensity $\mathbb{P}(Z(\boldsymbol{x}) > z + u \mid Z(\boldsymbol{x}) > u)$ and frequency $P(Z(\boldsymbol{x}) > u)$ at location $\boldsymbol{x}$ under GPD and binomial distributions, respectively, as in chapter 5 of Coles [2001].
  - Both models incorporate latent spatial process characterised by geographical and climatological covariates using a Gaussian process.
  - MCMC and spatial interpolation used for inference.
- temporal dependence reduced by declustering (keeping only highest of consecutive days exceeding threshold).
- Work in a "climate space" (rather than longitude/latitude), where the coordinates of $\boldsymbol{x}$ are defined by orographic and climatological measures.
- Both models are described by the three layer hierarchical model

$$p(\boldsymbol{\theta}|\boldsymbol{Z}(\boldsymbol{x})) \propto p_1(\boldsymbol{Z}(\boldsymbol{x})|\boldsymbol{\theta}_1)p_2(\boldsymbol{\theta}_1|\boldsymbol{\theta}_2)p_3(\boldsymbol{\theta}_2),$$

where the first layer models the data (with a GPD or Binomial), the second layer models the latent process (with a Gaussian process) and the third layer consists of the hyperparameters for the parameters θ_2 that drive the latent process (with θ_1 being the parameters of the GPD/binomial likelihoods).

- Best model amongst different covariate forms for $\phi = log(\sigma) and \xi$ were compared via DIC (also using long/lat vs climate space), ξ taken to be different for "mountainous" and "planes" regions, so fit for just **two separate regions** (with shared information) rather than estimated for each site (important as this is quite like clustering/grouping for parameter estimation improvement, although the clustering is via domain knowledge rather than data-driven).

A hierarchical max-stable spatial model for extreme precipitation

Leveraging Extremal Dependence to Better Characterize the 2021 Pacific Northwest Heatwave

References

Janet E. Heffernan and Jonathan A. Tawn. A conditional approach for multivariate extreme values (with discussion). *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 66(3): 497–546, July 2004. ISSN 1467-9868. doi: 10.1111/j.1467-9868.2004.02050.x. URL http://dx.doi.org/10.1111/j.1467-9868.2004.02050.x.

Caroline Keef, Ioannis Papastathopoulos, and Jonathan A. Tawn. Estimation of the conditional distribution of a multivariate variable given that one of its components is large: Additional constraints for the heffernan and tawn model. *Journal of Multivariate Analysis*, 115:396–404, March 2013a. ISSN 0047-259X. doi: 10.1016/j.jmva.2012.10.012. URL http://dx.doi.org/10.1016/j.jmva.2012.10.012.

Jonathan Tawn, Rob Shooter, Ross Towe, and Rob Lamb. Modelling spatial extreme events with environmental applications. *Spatial Statistics*, 28:39–58, 2018. ISSN 2211-6753. doi: https://doi.org/10.1016/j.spasta.2018.04.007. URL https://www.sciencedirect.com/science/article/pii/S2211675317302786. One world, one health.

JL Wadsworth and JA Tawn. Spatial conditional extremes. *Manuscript submitted for publication*, 2018.

Caroline Keef, Jonathan A. Tawn, and Rob Lamb. Estimating the probability of widespread flood events. *Environmetrics*, 24(1):13–21, 2013b. doi: https://doi.org/10.1002/env.2190. URL https://onlinelibrary.wiley.com/doi/abs/10.1002/env.2190.

Christian Rohrbeck and Jonathan A Tawn. Bayesian spatial clustering of extremal behavior for hydrological variables. *Journal of Computational and Graphical Statistics*, 30(1):91–105, 2021.

Elsa Bernard, Philippe Naveau, Mathieu Vrac, and Olivier Mestre. Clustering of maxima: Spatial dependencies among heavy rainfall in france. *Journal of climate*, 26(20):7929–7937, 2013.

Julie Carreau, P Naveau, and Luc Neppel. Partitioning into hazard subregions for regional peaks-over-threshold modeling of heavy precipitation. *Water Resources Research*, 53(5):4407–4426, 2017.

Daniel Cooley, Douglas Nychka, and Philippe Naveau. Bayesian spatial modeling of extreme precipitation return levels. *Journal of the American Statistical Association*, 102(479):824–840, 2007.

Edoardo Vignotto, Sebastian Engelke, and Jakob Zscheischler. Clustering bivariate dependencies of compound precipitation and wind extremes over great britain and ireland. *Weather and Climate Extremes*, 32:100318, 2021.

Debbie J Dupuis, Sebastian Engelke, and Luca Trapin. Modeling panels of extremes. *The Annals of Applied Statistics*, 17(1):498–517, 2023.

Miguel de Carvalho, Raphael Huser, and Rodrigo Rubio. Similarity-based clustering for patterns of extreme values. *Stat*, 12(1):e560, 2023. doi: https://doi.org/10.1002/sta4.560. URL https://onlinelibrary.wiley.com/doi/abs/10.1002/sta4.560.

Leonardo Bottolo, Guido Consonni, Petros Dellaportas, and Antonio Lijoi. Bayesian analysis of extreme values by mixture modeling. *Extremes*, 6:25–47, 2003.

Stuart Coles. *An introduction to statistical modeling of extreme values*. Springer series in statistics. Springer, Guildford, England, 2001 edition, November 2001.