Fast and flexible inference for spatial extremes

Peng Zhong, Scott A. Sisson, **Boris Béranger**

KAUST Statistics Workshop,
3 November 2025

UNSW Data Science Hub
uDASH

UNSW SYDNEY
Mathematics & Statistics

Introduction

Broad context

- Interest in the extremes of a stochastic process $X(s), s \in \mathcal{S}$.

E.g. $X(\cdot)$ *measures the amount rainfall at locations over Florida*

- **Goal:** Model the dependence structure in spatial extremes
- What characterises an extreme event? $\rightarrow$ Tailored approach
- Focus on asymptotic dependent processes: max-stable and r-Pareto.

In this talk

1. Establish theoretical conditions for max-stable and r-Pareto models to have a continuous exponent measure
2. Derive two new max-stable and r-Pareto models
3. Provide a fast inference methodology using spectral likelihoods

Modelling framework for max-stable and r-Pareto processes

Theoretical & methodological results

Simulation experiements

Analysis of extreme rainfall over Florida

Max-stable processes

Definition (Schlather, 2002)

A max-stable process with unit Fréchet margins can be characterized as

$$Z(\mathbf{s}) = \sup_{i=1}^{\infty} R_i W_i(\mathbf{s}), \ \mathbf{s} \in \mathcal{S},$$

where $R_1, R_2, \ldots$, are the points of a PPP on $(0, \infty)$ and $W_1(\mathbf{s}), W_2(\mathbf{s}), \ldots$, are independent copies of a stochastic processes $W(\mathbf{s})$ on $\mathcal{S}$ with unit mean.

The exponent measure restricted onto $\mathbb{R}_+^D$ is given by

$$\kappa\left([\mathbf{0}, \boldsymbol{x}]^c\right) = \int_0^\infty 1 - \Pr(\boldsymbol{W} \in [\mathbf{0}, \boldsymbol{x}r])\mathrm{d}r, \quad \mathbf{x} \in \Omega,$$

where $\boldsymbol{W} = (W(\mathbf{s}_1), \ldots, W(\mathbf{s}_D))^\top$ and $\Omega = \mathbb{R}_+^D \setminus \{\mathbf{0}\}$.

The distribution function can be expressed as

$$G(\boldsymbol{x}) = \exp\left\{-\kappa([0, \boldsymbol{x}]^c)\right\} = \exp\left\{-V(\boldsymbol{x})\right\}.$$

Max-stable processes

Let $B_D = \{1, \ldots, D\}$ and $B_k = \{b_1, \ldots, b_k\} \subset B_D$, where $b_1 < \cdots < b_k$.

Let $\Omega_{B_k} = \{\boldsymbol{x} \in \Omega : x_j = 0 \text{ if } j \notin B_k\}$ such that:

- $\partial\Omega = \{\Omega_{B_k}, \forall B_k \text{ and } k = 1, \ldots, D-1\}$ represents the boundaries of Ω,
- $\Omega^\circ = \Omega \setminus \partial\Omega$ denotes the Interior of Ω.

Important

Depending on the choice of $\boldsymbol{W}$, the exponent measure κ **can put mass on both $\partial\Omega$ and Ω°** with the intensity function on each subspace Ω_{B_k}

$$\lim_{x_i \to 0, i \notin B_k} -V_{B_k}(\boldsymbol{x}), \quad V_{B_k} = \frac{\partial^k V}{\partial x_{b_1} \ldots \partial x_{b_k}}.$$

On Ω°, it can be expressed as $\kappa(\boldsymbol{x}) = -V_{B_D}(\boldsymbol{x})$, where the function κ is referred to as the intensity function of the max-stable process.

Max-stable processes - Inference

Full likelihood: intractable!

Composite likelihood: Popular but still limited.

Stephenson-Tawn likelihood: Can be biased, moderate dimensions.

Spectral likelihood (Coles & Tawn, 1991)

If data $\in$ MDA($\boldsymbol{Z}$) then can be approximately treated as points of a PPP with measure $\kappa(\cdot)$. For a model with parameter $\boldsymbol{\theta}$, the log-likelihood is

$$\ell_A(\boldsymbol{\theta}; \mathbf{x}_1, \ldots, \mathbf{x}_n) \propto \sum_{i \in \{m : \|\mathbf{x}_m\|_1 > u\}} \log \kappa\left(\boldsymbol{x}_i; \boldsymbol{\theta}\right).$$

for some high enough threshold u.

This requires convergence of:

a) X to the max-stable process Z by taking pointwise maxima.

b) X to the Poisson point process.

The fact that κ can put mass on $\partial\Omega$ hinders the convergence of $X \Longrightarrow$ bias.

r-Pareto processes

Definition (Dombry & Ribatet, 2015)

Assuming the process X with unit Pareto margins satisfying $\lim_{u\to\infty} u \Pr(X/u \in B) = \kappa(B), \forall B \subset C^+(\mathcal{S})$, then the limiting process

$$\tilde{Z}(\mathbf{s}) = \lim_{u\to\infty} \frac{X(\mathbf{s})}{u} | r\left(\{X(\mathbf{s}), \mathbf{s} \in \mathcal{S}\}\right) > u,$$

defines a simple r-Pareto process on $\mathcal{A}_r = \{f \in C^+(\mathcal{S}) : r(f) > 1\}$ with probability measure $\kappa(\cdot \cap \mathcal{A}_r)/\kappa(\mathcal{A}_r)$.

The finite dimensional density is therefore

$$\frac{\kappa(\mathbf{x})}{\kappa(\mathcal{A}_r^D)}, \quad \mathbf{x} \in \mathcal{A}_r^D,$$

where κ is the intensity function and $\mathcal{A}_r^D$ is the set $\mathcal{A}_r$ restricted to D dimensions.

r-Pareto processes - Inference

The log-likelihood is thus

$$\ell_{rP}(\boldsymbol{\theta}; \mathbf{x}_1, \dots, \mathbf{x}_n) = \sum_{i \in \{m : r(\mathbf{x}_m) > u\}} \log\left(\frac{\kappa(\mathbf{z}_i; \boldsymbol{\theta})}{\kappa(\mathcal{A}_r; \boldsymbol{\theta})}\right),$$

where $\mathbf{z}_i = \mathbf{x}_i / u$ represent the realizations of the r-Pareto process.

Important

- $\kappa(\mathcal{A}_r; \boldsymbol{\theta})$ involves integration over $\mathbb{R}_+^D$, $\Longrightarrow$ intractability
- de Fondeville & Davison (2018):
  - ⋆ Simplifications for specific choices of $r(\cdot)$.
  - ⋆ Score matching.
- $r(\mathbf{x}) = \|\mathbf{x}\|_1 \Longrightarrow$ spectral likelihood.
- If the exponent measure κ has discontinuities (presence of mass on $\partial\mathcal{A}_r^D$),

$\Longrightarrow$ Inference requires evaluation of $-V_{B_k}(\boldsymbol{x})$.

  - ⋆ Restriction to the Brown-Resnick models

Modelling framework for max-stable and r-Pareto processes

Theoretical & methodological results

Simulation experiements

Analysis of extreme rainfall over Florida

Ensuring continuous exponent measures

Theorem 1 (Zhong, Sisson & Béranger, 2025)

Consider the max-stable process $\{Z(\mathbf{s}), \mathbf{s} \in \mathcal{S}\}$ defined at D locations and assume the partial derivatives of the function V exist.

The intensity function on $\partial\Omega$ is zero almost everywhere if and only if the conditional probability of $\boldsymbol{W}$ satisfies

$$\Pr(\boldsymbol{W}_{\bar{B}_k} = \mathbf{0}_{D-k} \mid \boldsymbol{W}_{B_k} = \boldsymbol{x}_{B_k}) = 0, \ \forall\, k \in \{1, \ldots, D-1\},$$

where $\boldsymbol{x}_{B_k} > \mathbf{0}_k$.

Interpretation: the behavior of V_{B_k} on $\partial\Omega$ is determined by the behavior of the process W on its lower-end boundary at zero.

Ensuring continuous exponent measures

Brown-Resnick: $W = \exp\left(\tilde{W} - \frac{\sigma^2}{2}\right)$ with $\tilde{W}$ a centered Gaussian process

$$\left\{\boldsymbol{W}_{\bar{B}_k} = \boldsymbol{0}_{D-k} \mid \boldsymbol{W}_{B_k} = \boldsymbol{x}_{B_k}\right\}$$
$$\Longleftrightarrow \left\{\tilde{\boldsymbol{W}}_{\bar{B}_k} = -\infty_{D-k} | \tilde{\boldsymbol{W}}_{B_k} = \log(\boldsymbol{x}_{B_k}) + \sigma^2/2\right\}.$$

The conditional distribution $\tilde{\boldsymbol{W}}_{\bar{B}_k} | \tilde{\boldsymbol{W}}_{B_k}$ is Gaussian $\Longrightarrow$ Condition satisfied

skew extremal-t : $W = \max(\tilde{W}^\nu, 0)$ with $\tilde{W}$ a skew-normal process, $\nu > 0$.

$$\{\boldsymbol{W}_{\bar{B}_k} = \boldsymbol{0}_{D-k} | \boldsymbol{W}_{B_k} = r\boldsymbol{x}_{B_k}\}$$
$$\Longleftrightarrow \left\{\tilde{\boldsymbol{W}}_{\bar{B}_k} \leq \boldsymbol{0}_{D-k} | \tilde{\boldsymbol{W}}_{B_k} = (r\boldsymbol{x}_{B_k})^{1/\nu}\right\},$$

This is not a null event $\Longrightarrow$ Condition NOT satisfied

Extending current classes of max-stable and r-Pareto models

Theorem 2 (Zhong, Sisson & Béranger, 2025)

a) skewed Brown-Resnick: Let $W(\mathbf{s}) = \exp\{Y(\mathbf{s}) - a(\mathbf{s})\}$ where $Y(\mathbf{s})$ is a centred skew-normal process with scale matrix Σ with slant parameter $\boldsymbol{\alpha}$, and $a(\mathbf{s}) = \log \mathbb{E}\left[\exp\{Y(\mathbf{s})\}\right]$.

b) truncated extremal-t: Let $W(\mathbf{s}) = \tilde{Y}(\mathbf{s})^\nu / a(\mathbf{s})$, with $\nu > 0$, $\tilde{Y}(\mathbf{s}) = Y(\mathbf{s}) | Y(\mathbf{s}) > 0$, where $Y(\mathbf{s})$ is a centred Gaussian process with unit variances and $a(\mathbf{s}) = \mathbb{E}\left[\tilde{Y}(\mathbf{s})^\nu\right]$.

$\Longrightarrow$ Both models have no mass on $\partial\Omega$.

Comments:

- The sBR model has a non-stationary dependence structure.
- The intensity of the truncated extremal-t is somewhat difficult to compute...
- Removal of the mass on $\partial\Omega$ increases the dependence strength

Extending current classes of max-stable and *r*-Pareto models

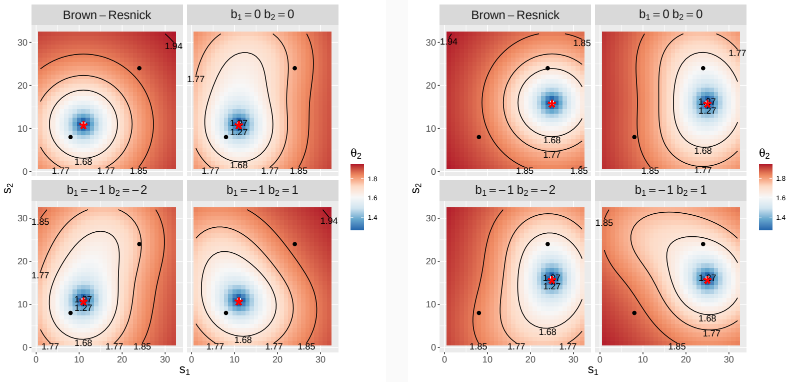

Figure 1: Bivariate extremal coefficient for the Brown-Resnick model and skewed Brown-Resnick model where $\eta_i = \sum_{j=1}^{2} b_j K_j(\boldsymbol{s}_i) + 0.1\text{sgn}(s_{2,i} \geq 16), i = 1, \ldots, D$ using $(b_1, b_2) = (0, 0), (-1, -2)$ and $(-1, -1)$. Black dots denote the kernel centres $\boldsymbol{s}_1^\star, \boldsymbol{s}_2^\star$. A red star indicates the reference point.

Improved inference for r-Pareto models

Where does the idea come from?

[Dombry, Legrand & Opitz (2024)] Using rejection sampling, one can generate samples from a r-Pareto process with risk functional r_2 from samples of a r-Pareto process associated with risk functional r_1 as long as $Mr_1(\cdot) \geq r_2(\cdot), M > 0$.

Focus: Observations $i \in \{m : r(\mathbf{x}_m) > u\}$

Proposal: use the likelihood of the L_1-Pareto process to make inference about any r-Pareto process with a different risk functional by choosing a high threshold $u > M$.

This particularly applies to L_p norms, $p > 1$, since L_p bounds L_1 for finite p.

$\Longrightarrow \|\cdot\|_1 \leq D^{1-1/p}\|\cdot\|_p,\ p > 1$

$\Longrightarrow$ choose $u > D^{1-1/p}, p > 1$, to infer the L_p-Pareto process.

Benefit: Avoids to compute the normalising constant!!

Modelling framework for max-stable and r-Pareto processes

Theoretical & methodological results

Simulation experiements

Analysis of extreme rainfall over Florida

Simulation algorithm for r-Pareto processes

Current algorithm (Dombry et al., 2024)

Let $r_0(\cdot) = \|\cdot\|_1$
$\hookrightarrow$ samples can be obtained as $\tilde{R}\boldsymbol{Y}/\|\boldsymbol{Y}\|_1$, where $\tilde{R} \sim$ Pareto(1).

For another risk functional $r_1(\cdot)$ be such that $r_1(\cdot) \leq Mr_0(\cdot)$
$\hookrightarrow$ samples can then be obtained by sampling $\tilde{\boldsymbol{Z}}$ from the process with risk functional $r_0(\cdot)$, and accepting a sample as $\tilde{\boldsymbol{Z}}/M$ when $r_1(\tilde{\boldsymbol{Z}}) \geq M$.

Comments:

- $M = 1$ when $r_1 = \|\cdot\|_\infty$
- $M = D^{p-2}$ when $r_1 = \|\cdot\|_p, p > 1$.
- Larger M implies lower acceptance probabilities.

Simulation algorithm for r-Pareto processes

Theorem 3 (Zhong, Sisson & Béranger, 2025)

Let $\tilde{\boldsymbol{Z}}^{(r_0)}$ be a r-Pareto process at D locations with $r_0(\cdot) = \|\cdot\|_1$.

Let $r(\cdot)$ be any convex risk functional, homogeneous of order 1, such that $r(\mathbf{0}) = 0$ and $r(\boldsymbol{e}_i) = 1/c_i > 0, i = 1, \ldots, D$, where $\boldsymbol{e}_i$ are standard basis vectors of dimension D.

The r-Pareto process with risk functional $r(\cdot)$ can be simulated as $\tilde{\boldsymbol{Z}}^{(r)} \stackrel{d}{=} c_0 \tilde{\boldsymbol{Z}}^{(r_0)} \mid r\left(\tilde{\boldsymbol{Z}}^{(r_0)}\right) > 1/c_0$, where $c_0 = \min\{\min_{i=1}^{D} c_i, 1\}$.

Comments:

- Considering $r(\cdot) = \|\cdot\|_p,\ p > 1$ gives $1/c_0 = 1$ instead of $M = D^{p-2}$ in Dombry et al (2024).
- $r(\boldsymbol{x}) = \sum_{i=1}^{D} m_i x_i, m_i \in [0, \infty], i = 1 \ldots, D$, a linear convex combination, produces $c_0 = \min\{\min_{i=1}^{D} 1/m_i, 1\}$.

Simulation algorithm for r-Pareto processes

Table 1: Percentage of samples from 10^5 Brown-Resnick r-Pareto processes replicates with L_1 risk functional that fall into the acceptance region for $r(\cdot)$ being $L_p, p = 2, 3, 5, 10$, using the current and proposed methods (resp. left and right number in each cell). Dimension is $D = 4, 16, 64, 100$ and dependence structure is defined via a power-law semivariogram $\gamma(h) = h/\lambda$, $\lambda = 2$, on the grid $[1, \sqrt{D}]^2$.

L_p/D	$D = 4$	$D = 16$	$D = 64$	$D = 100$
$p = 2$	59/59	37/37	25/25	22/22
$p = 3$	13/52	1.80/29	0.27/18	0.19/16
$p = 5$	0.76/49	>0.01/26	0.00/16	0.00/14
$p = 10$	>0.01/47	0.00/25	0.00/15	0.00/13

Spectral likelihoods vs score matching

Setup:

- Generate $n = 2,000$ obs from the skewed Brown-Resnick model on a 15×15 grid ($D = 225$).
- Power-law semivariogram $\gamma(h) = (h/\lambda)^{\vartheta}$ with range $\lambda = 5, 10$ and smoothness $\vartheta = 1, 1.5$.
- Skewness represented through spline functions with 2 Gaussian kernel basis functions $(b_1, b_2) = (0, 0), (-1, -2), (-1, 1)$.
- L_1 and L_3 risk functionals.
- An observation is considered extreme when exceeding the 95% empirical quantile of $r(\boldsymbol{X}_1), \ldots, r(\boldsymbol{X}_n)$.
- 300 replicates.

Spectral likelihoods vs score matching

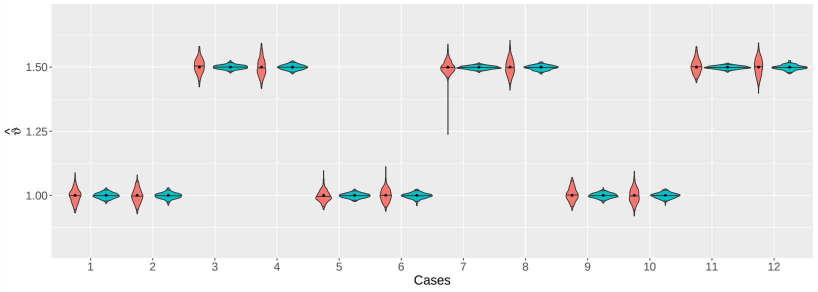

Figure 2: Violin plots for score matching (red) and spectral likelihood (blue) estimates of ϑ for the skewed Brown-Resnick r-Pareto process with L_3 norm risk functional. Black dots indicate the parameter true values.

- The spectral likelihood provides unbiased, low variability estimates.
- The score matching produces unbiased but more variable estimates.

Spectral likelihoods vs score matching

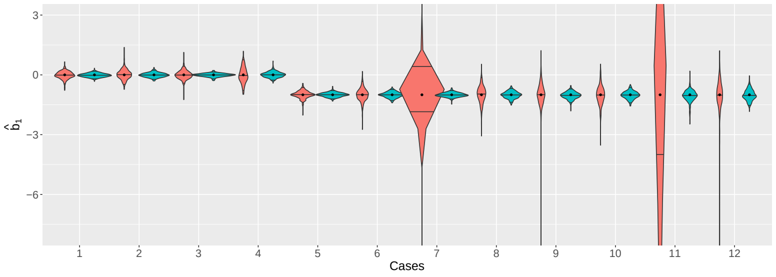

Figure 3: Violin plots for score matching (red) and spectral likelihood (blue) estimates of b_1 for the skewed Brown-Resnick r-Pareto process with L_3 norm risk functional. Black dots indicate the parameter true values.

- Score matching estimates can become numerically unstable (cases 7–12).
- Spectral likelihood is ~ 5 times faster than the score matching approach (141 versus 704 seconds on average using 3 CPU cores).

Modelling framework for max-stable and r-Pareto processes

Theoretical & methodological results

Simulation experiements

Analysis of extreme rainfall over Florida

Analysis of extreme rainfall over Florida

Data:

- Location: Tampa Bay area, Florida. Regular 2km grid with $4,449$ spatial observations.

- Measurements: radar images recorded at 15 minute intervals between 1995–2019 during the wet season (June–September). Total $n = 139,881$ images.

- Smaller version of the dataset analysed in de Fondeville & Davison (2018).

- Risk functions:

  $\rightarrow$ L_∞ norm: defines extremes events as locally intense rainfall events at any location within the region

  $\rightarrow$ L_1 norm selects events with high cumulative rainfall over the whole region.

Analysis of extreme rainfall over Florida

Modelling:

- Brown-Resnick (BR) and skewed Brown-Resnick (sBR) with anisotropic semivariogram.
- Skewness of sBR expressed using 4 kernels.
- Fitting using score matching and spectral likelihood.

Outcomes:

- Brown-Resnick:

  $\rightarrow$ Spectral likelihood and score matching provide consistent estimates.

  $\rightarrow$ Spectral likelihood is 80% (L_1 norm) and 18% (L_∞ norm) faster.

- Brown-Resnick vs skewed Brown-Resnick:

  $\rightarrow$ AIC favours the skewed Brown-Resnick for both L_1 and L_∞ norms.

Analysis of extreme rainfall over Florida

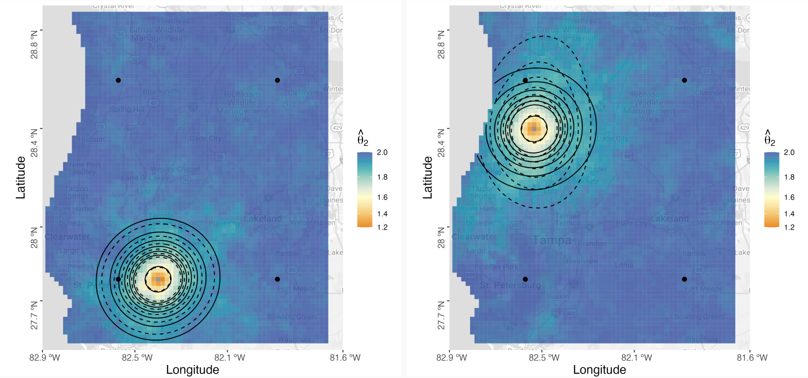

Figure 4: Maps of bivariate empirical extremal coefficients (shading) with respect to two different reference points, and contours of the extremal coefficient of the fitted sBR (dashed line) and BR (solid line) r-Pareto models with L_∞ norm risk functional. Black dots denote the kernel centres used in the sBR model.

Conclusion

- Established condition ensuring the intensity function of a max-stable process only places mass on Ω°;

  $\rightarrow$ No discontinuities in the associated exponent measure;

  $\rightarrow$ Simplifying the evaluation of the density of the r-Pareto process.

- Likelihood-based inference can be successfully implemented via the spectral likelihood.

- Two new models: skewed Brown-Resnick and truncated Extremal-t.

- Improved rejection sampling algorithm for r-Pareto processes.

THANK YOU

https://arxiv.org/pdf/2407.13958

B.Beranger@unsw.edu.au