Modelling extremal dependence of a 3-dimensional oceanographic dataset via a Bayesian geometric approach

Ryan Campbell

Lancaster University

Ex-CRG Workshop
24 October 2023

Papastathopoulos et al. (2023)

THE UNIVERSITY OF EDINBURGH

THE UNIVERSITY
of EDINBURGH

KAUST
King Abdullah University of
Science and Technology

Limit sets and gauge functions

- $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n \overset{\text{iid}}{\sim} f$, d-dimensional, Laplace$(0, 1)$ margins.
- Scaled sample clouds $\left\{ \frac{\boldsymbol{X}_1}{\log(n/2)}, \ldots, \frac{\boldsymbol{X}_n}{\log(n/2)} \right\}$ converge onto a **limit set**,

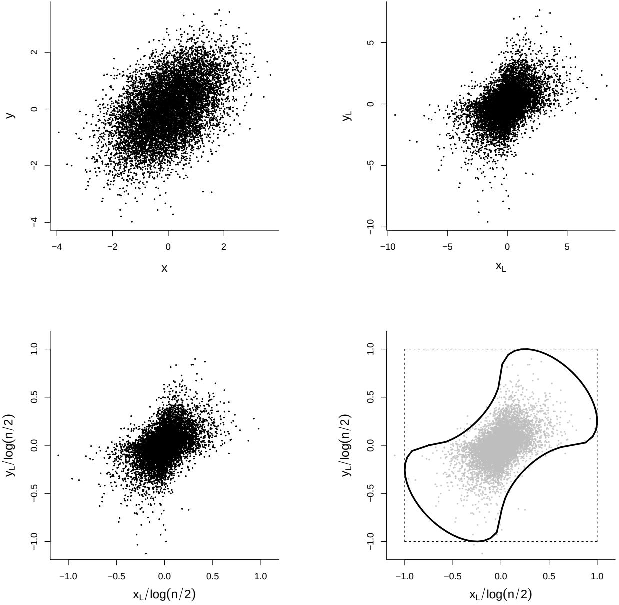
$$G := \left\{ \boldsymbol{x} \in \mathbb{R}^d \;\middle|\; g(\boldsymbol{x}) \leq 1 \right\}$$

  as $n \to \infty$ (Balkema et al., 2010).
- The **gauge function**, $g : \mathbb{R}^d \longrightarrow \mathbb{R}$, is 1-homogeneous and is obtained through:

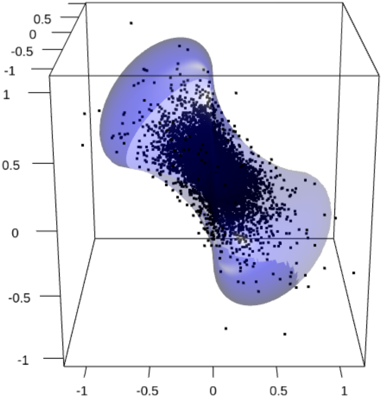
$$g(\boldsymbol{x}) = \lim_{t \to \infty} \frac{-\log f(t\boldsymbol{x})}{t}$$

  (Nolde, 2014; Nolde and Wadsworth, 2021)

Example: Gaussian, $\rho = 0.5$



$d = 3$ Gaussian, $\rho_{12} = -0.5$, $\rho_{13} = -0.5$, $\rho_{23} = 0.5$

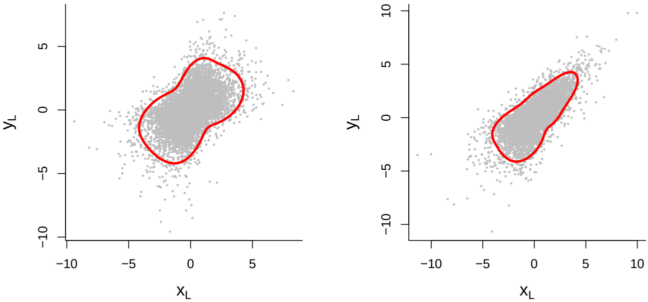


Extremal inference with gauge functions

- Consider

$$\begin{aligned} R &= \|\boldsymbol{X}\|_2 \\ \boldsymbol{W} &= \boldsymbol{X}/\|\boldsymbol{X}\|_2 \\ r_q(\boldsymbol{w}) &= q^{\text{th}} \text{ quantile of } R \text{ given angle } \boldsymbol{W} = \boldsymbol{w} \\ \text{i.e.,} \quad & \Pr\left(R \leq r_q(\boldsymbol{w}) | \boldsymbol{W} = \boldsymbol{w}\right) = q \end{aligned}$$

- Interested in when R is large, or $R > r_q(\boldsymbol{W})$

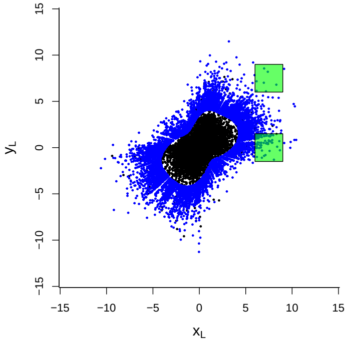


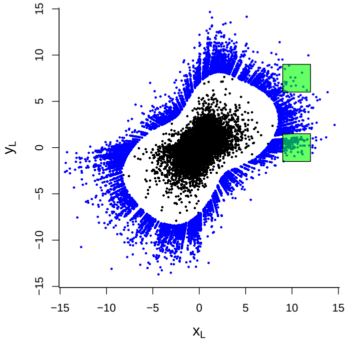
Parametric Approach (Wadsworth and Campbell, 2022)

- Wadsworth and Campbell (2022) establish

$$R \mid \{\boldsymbol{W} = \boldsymbol{w}, R > r_q(\boldsymbol{w})\} \sim \text{truncatedGamma}(d, g(\boldsymbol{w}; \boldsymbol{\theta}))$$

- Empirical distribution for $\boldsymbol{W} \mid R > r_q(\boldsymbol{W})$





Bayesian approach

- Here, we fit the model

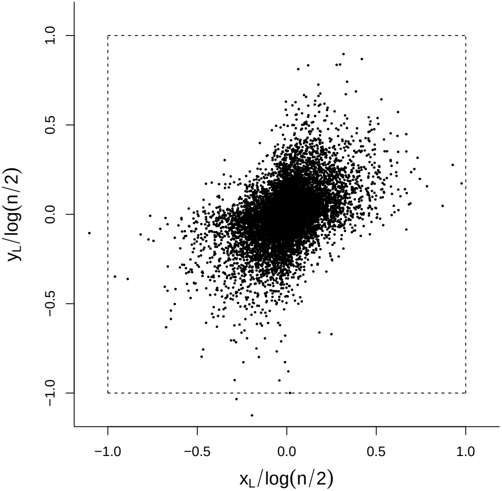
$$R - r_q(\boldsymbol{w}) \mid \{\boldsymbol{W} = \boldsymbol{w}, R > r_q(\boldsymbol{w})\} \sim \text{Exponential}(g(\boldsymbol{w}; \eta, Q^{-1}))$$

where

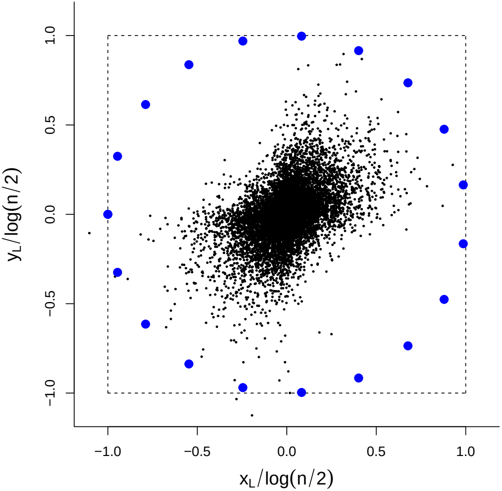
$$g(\boldsymbol{w}; \eta, Q^{-1}) = \exp\left\{\eta + \sum_{i=1}^{N} \phi_i h_i(\boldsymbol{w})\right\} \quad ; \quad \boldsymbol{\phi} \sim \text{MVN}_N(0, Q^{-1})$$

- Q: sparse $N \times N$ precision matrix
- N: angular mesh length
- $g(\boldsymbol{x}) = g(r\boldsymbol{w}) = r\, g(\boldsymbol{w}; \eta, Q^{-1})$

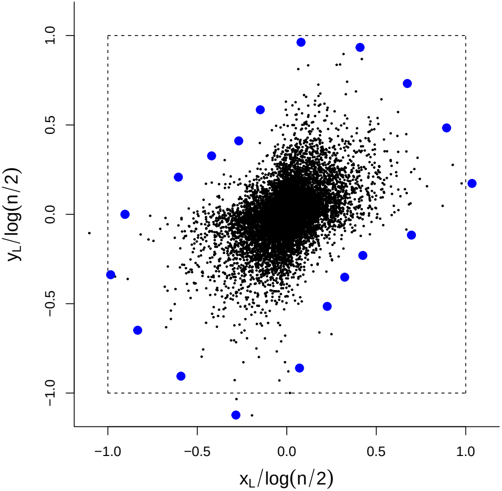
Example: Gaussian, $\rho = 0.5$



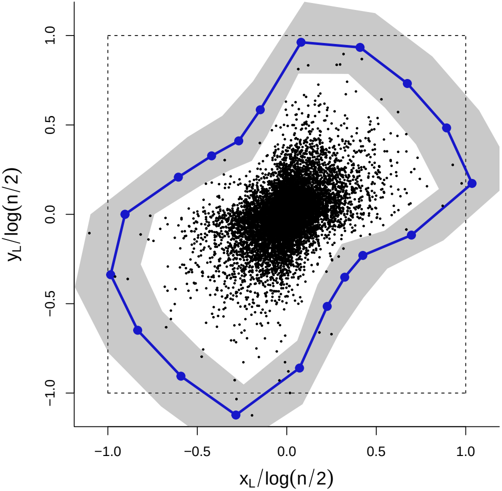
Example: Gaussian, $\rho = 0.5$, $N = 20$



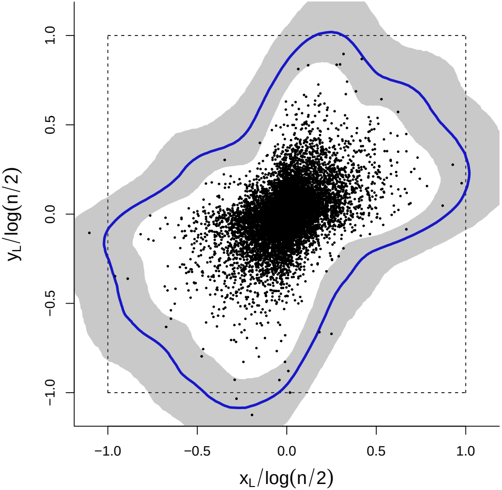
Example: Gaussian, $\rho = 0.5$, $N = 20$



Example: Gaussian, $\rho = 0.5$, $N = 20$

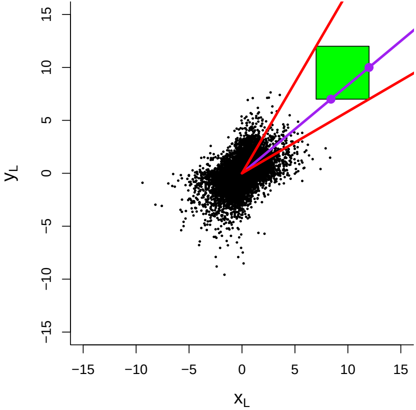


Example: Gaussian, $\rho = 0.5$, $N = 5,000$



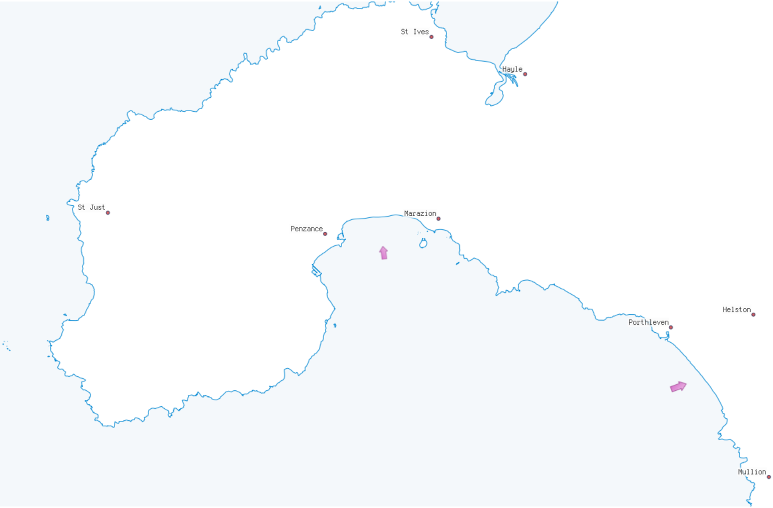
Probability Estimation

$$
\begin{aligned}
Pr\left(\boldsymbol{X} \in B\right) &= \Pr\left(r_{\inf}(\boldsymbol{W}) \leq R \leq r_{\sup}(\boldsymbol{W}), \boldsymbol{W} \in \mathcal{S}_B\right) \\
&= \Pr\left(\boldsymbol{W} \in \mathcal{S}_B\right) \int\limits_{\mathcal{S}_B} \left[e^{-(r_{\inf}(\boldsymbol{w}) - r_0(\boldsymbol{w}))g(\boldsymbol{w})} - e^{-\left(r_{\sup}(\boldsymbol{w}) - r_0(\boldsymbol{w})\right)g(\boldsymbol{w})} \right] d\boldsymbol{w}
\end{aligned}
$$



Newlyn wave data example ($d = 3$)

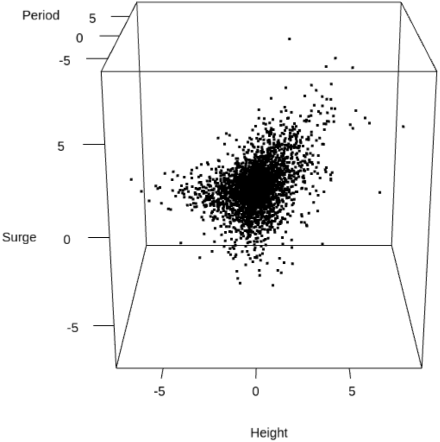
Newlyn wave data example ($d = 3$)



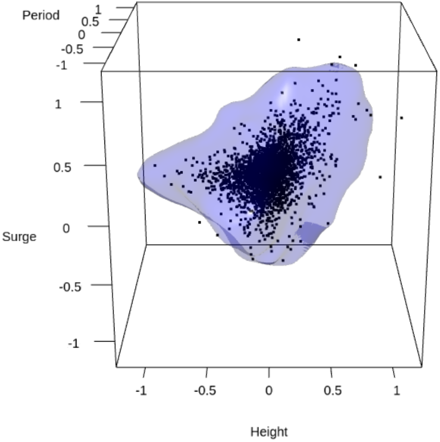
Newlyn wave data example ($d = 3$)

- $n = 2,894$, $d = 3$
  - H, wave height (meters, m)
  - P, wave period (seconds, s)
  - S, wave surge (meters, m)

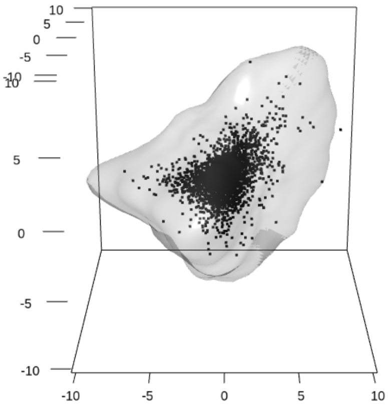
Newlyn wave data example ($d = 3$)



Newlyn wave data example ($d = 3$)



Newlyn wave data example ($d = 3$)



Newlyn wave data example ($d = 3$)

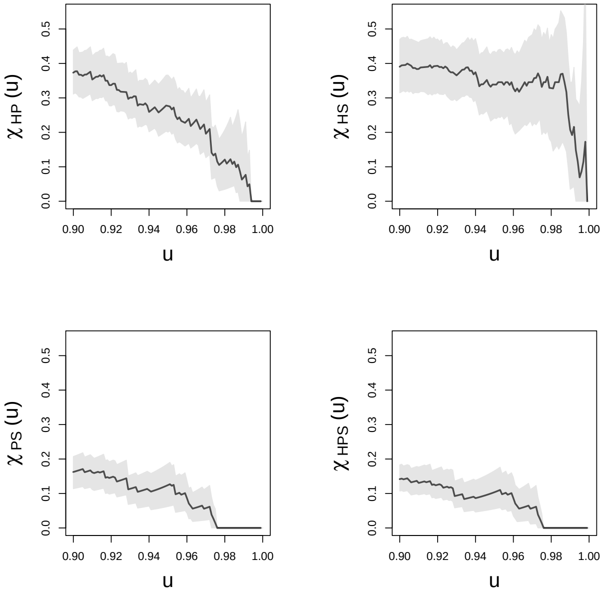
- Interested in

$$\chi(u) = \left(\frac{1}{1-u}\right) \Pr\left[F_L\left(X_j\right) > u \,\forall\, j \in A \subseteq \{H, P, S\}\right]$$

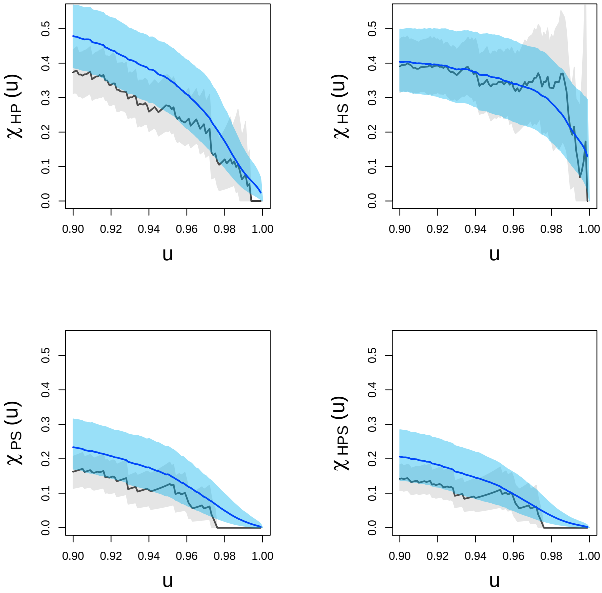
  for u close to 1

  - $\chi_{HP}(u) = \Pr\left(F_L(X_H) > u | F_L(X_P) > u\right)$
  - $\chi_{HS}(u) = \Pr\left(F_L(X_H) > u | F_L(X_S) > u\right)$
  - $\chi_{PS}(u) = \Pr\left(F_L(X_P) > u | F_L(X_S) > u\right)$
  - $\chi_{HPS}(u) = \Pr\left(F_L(X_H) > u, F_L(X_P) > u | F_L(X_S) > u\right)$

Newlyn wave data example ($d = 3$)



Newlyn wave data example ($d = 3$)



Newlyn wave data example ($d = 3$)

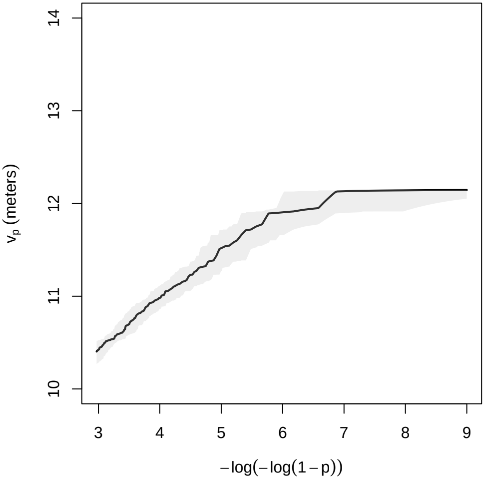
- The sea-wall *overlapping discharge rate* ($m^3 s^{-1} m^{-1}$)

$$Q(v; X_H, X_P, X_S) = a_1 X_S X_P \exp\left\{ \frac{a_2 (v - X_S - l)}{(X_P {X_H^*}^{1/2})} \right\}$$

  - v: sea-wall height (m)
  - $a_1 = 0.25$, $a_2 = 26$: sea-wall design features
  - $l = 4.3$: tidal level relative to the seabed
- Industrial design standard of $0.002 m^3 s^{-1} m^{-1}$
- Want to estimate v_p where $p = \Pr(V > v_p)$.
- Coles and Tawn (1994); Bortot et al. (2000); Wadsworth and Campbell (2022)

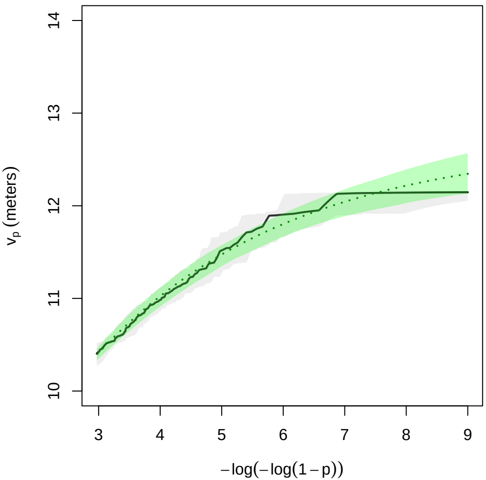
Newlyn wave data example ($d = 3$)

- p vs. $v_p := F_V^{\leftarrow}(1 - p)$



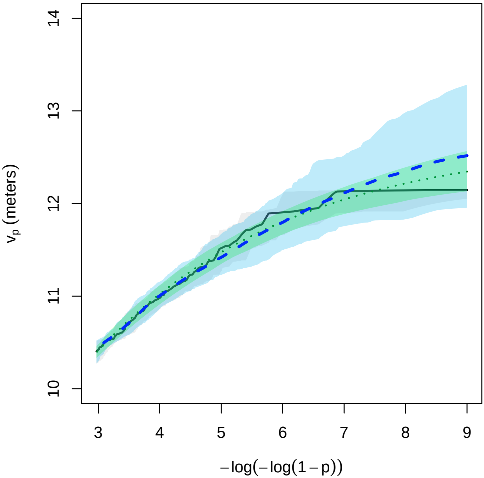
Newlyn wave data example ($d = 3$)

- p vs. $v_p := F_V^{\leftarrow}(1 - p)$



Newlyn wave data example ($d = 3$)

- p vs. $v_p := F_V^{\leftarrow}(1-p)$



Our Bayesian Method

- Using INLA, we can sample posterior observations from $r_q(\boldsymbol{w})$, $f_W(\boldsymbol{w})$, $g(\boldsymbol{w})$, and "return level curves".
- Backed up by rigorous theory.
- Performs well on simulation studies for a wide range of dependence structures.
- Allows for flexible modelling of multivariate extreme value problems.

Bibliography

A. A. Balkema, P. Embrechts, and N. Nolde. Meta densities and the shape of their sample clouds. *J. Multivar. Anal.*, 101(7):1738–1754, aug 2010. ISSN 0047-259X. doi: 10.1016/j.jmva.2010.02.010. URL https://doi.org/10.1016/j.jmva.2010.02.010.

P. Bortot, S. Coles, and J. Tawn. The multivariate Gaussian tail model: An application to oceanographic data. *JRSSC*, 49(1):31–049, 2000.

S. G. Coles and J. A. Tawn. Statistical methods for multivariate extremes: an application to structural design. *JRSSC*, 43(1):1–31, 1994.

N. Nolde. Geometric interpretation of the residual dependence coefficient. *Journal of Multivariate Analysis*, 123:85–95, 2014. ISSN 0047-259X.

N. Nolde and J. Wadsworth. Linking representations for multivariate extremes via a limit set. *Advances in Applied Probability*, 2021.

I. Papastathopoulos, L. de Monte, R. Campbell, and H. Rue. Statistical inference for radially-stable generalized pareto distributions and return level-sets in geometric extremes, 2023.

J. Wadsworth and R. Campbell. Statistical inference for multivariate extremes via a geometric approach. *arXiv preprint arXiv:2208.14951*, 2022.