Generative modelling of multivariate geometric extremes using normalising flows

Jordan Richards[1]

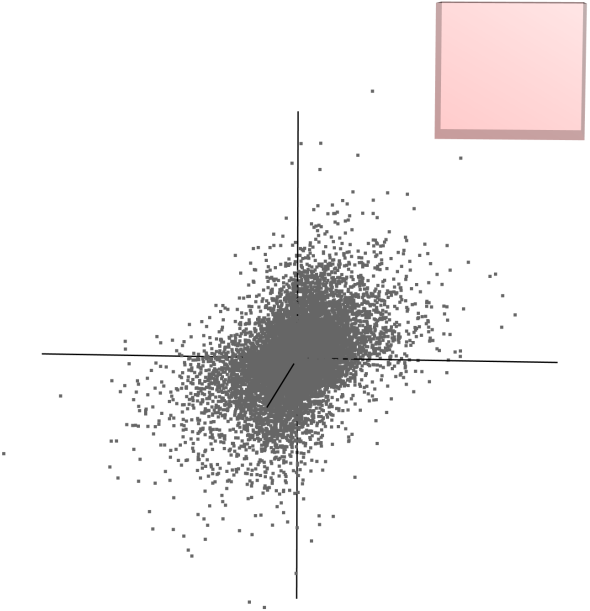
Joint work with Lambert De Monte[1], Raphaël Huser[2], Ioannis Papastathopoulos[1]

THE UNIVERSITY *of* EDINBURGH
School of Mathematics

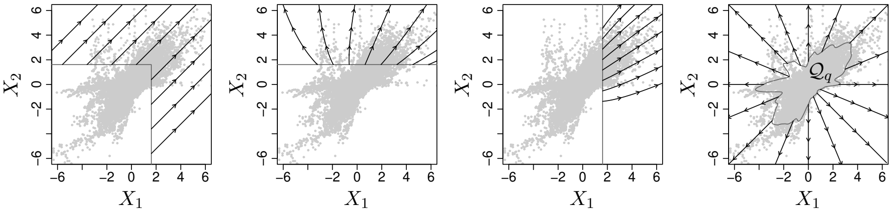
جامعة الملك عبدالله للعلوم والتقنية
King Abdullah University of Science and Technology

[1] University of Edinburgh and Maxwell Institute, [2] King Abdullah University of Science and Technology

Very broad overview

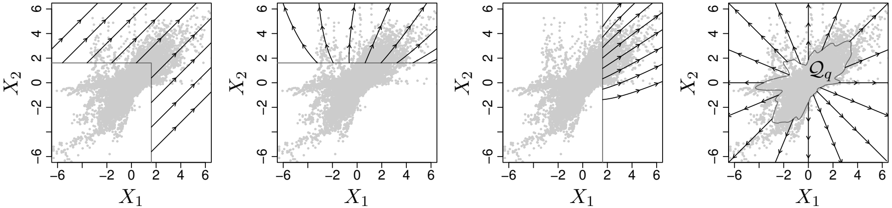


Geometric multivariate EVT: Motivation



Directions along which MEVT frameworks allow extrapolation to tail regions: (a) MRV, (b) and (c) conditional extremes, (d) geometric extremes.

Geometric multivariate EVT: Motivation



Directions along which MEVT frameworks allow extrapolation to tail regions: (a) MRV, (b) and (c) conditional extremes, (d) geometric extremes.

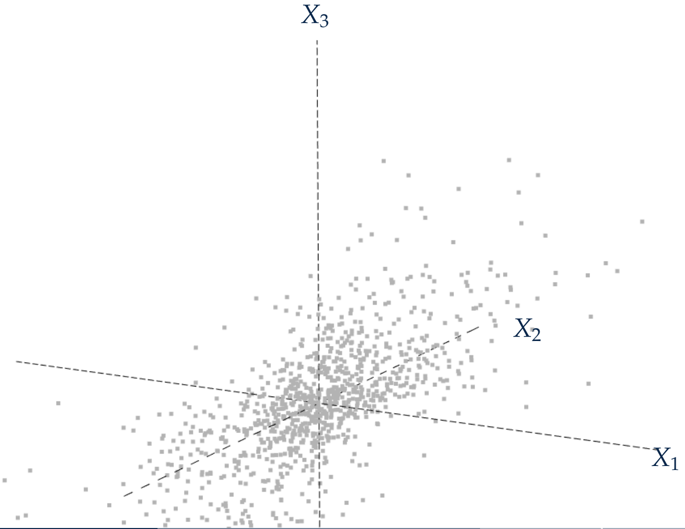
- Aim to extend probability estimation via a semi-parametric, geometric approach to multivariate extremes to higher-dimensional settings.
- Leverage theoretical links between the geometry of (starshaped set) parameters to define a range of parsimonious to flexible models.
- Use the generative framework of normalising flows to enable fast sampling and probability estimation.

Presentation Overview

1. Geometric extreme value theory
2. Statistical inference
3. Simulation study
4. An application to low and high wind speeds

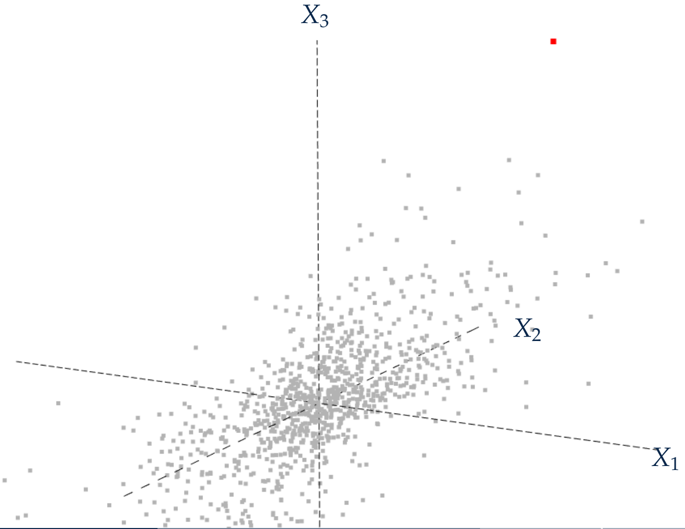
Very broad overview – A directional statistics approach

Let $\boldsymbol{X}_1, \boldsymbol{X}_2, \ldots \in \mathbb{R}^d$ be iid draws from $\mathbb{P}_{\boldsymbol{X}}$ with standard Laplace marginal distributions



Very broad overview – A directional statistics approach

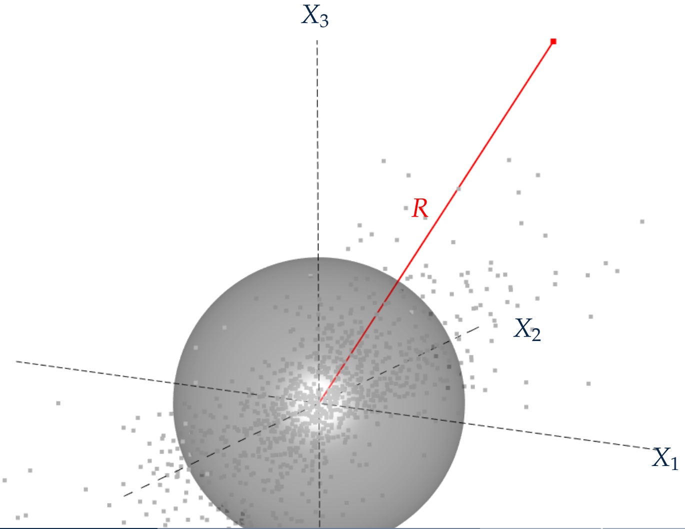
Let $\boldsymbol{X}_1, \boldsymbol{X}_2, \ldots \in \mathbb{R}^d$ be iid draws from $\mathbb{P}_{\boldsymbol{X}}$ with standard Laplace marginal distributions



Very broad overview – A directional statistics approach

Let $\boldsymbol{X}_1, \boldsymbol{X}_2, \ldots \in \mathbb{R}^d$ be iid draws from $\mathbb{P}_{\boldsymbol{X}}$ with standard Laplace marginal distributions, and

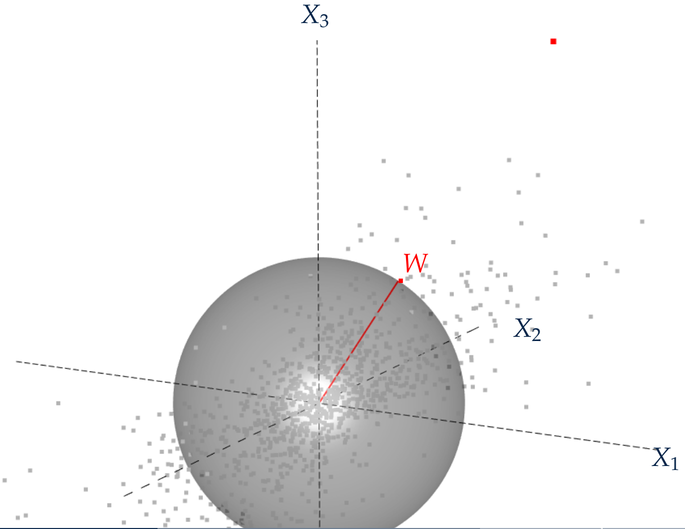
$$R := \|\boldsymbol{X}\|_2 > 0, \qquad \boldsymbol{W} := \boldsymbol{X}/R \in \mathbb{S}^{d-1}.$$



Very broad overview – A directional statistics approach

Let $\boldsymbol{X}_1, \boldsymbol{X}_2, \ldots \in \mathbb{R}^d$ be iid draws from $\mathbb{P}_{\boldsymbol{X}}$ with standard Laplace marginal distributions, and

$$R := \|\boldsymbol{X}\|_2 > 0, \qquad \boldsymbol{W} := \boldsymbol{X}/R \in \mathbb{S}^{d-1}.$$

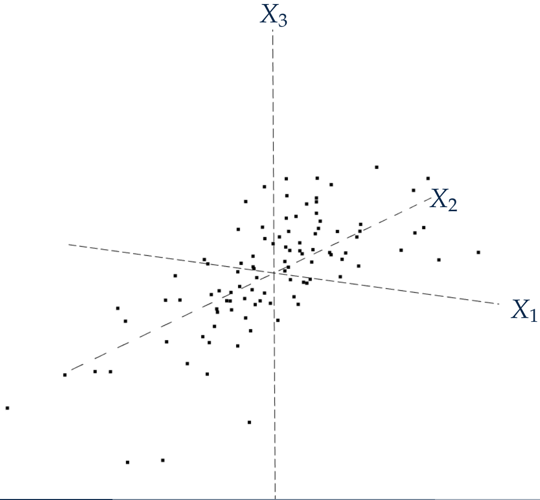


Limit sets

Let $X_1, X_2, \ldots \in \mathbb{R}^d$ be iid draws from $\mathbb{P}_X$ with standard Laplace marginal distributions, and

$$N_n := \left\{ \frac{X_1}{\log(n)}, \ldots, \frac{X_n}{\log(n)} \right\}$$

$n = 100$

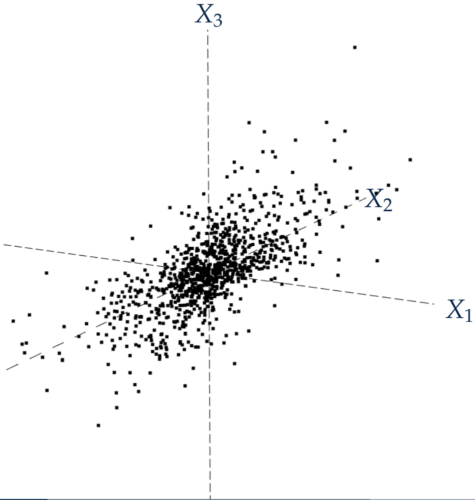


Limit sets

Let $\boldsymbol{X}_1, \boldsymbol{X}_2, \ldots \in \mathbb{R}^d$ be iid draws from $\mathbb{P}_{\boldsymbol{X}}$ with standard Laplace marginal distributions, and

$$N_n := \left\{ \frac{\boldsymbol{X}_1}{\log(n)}, \ldots, \frac{\boldsymbol{X}_n}{\log(n)} \right\}$$

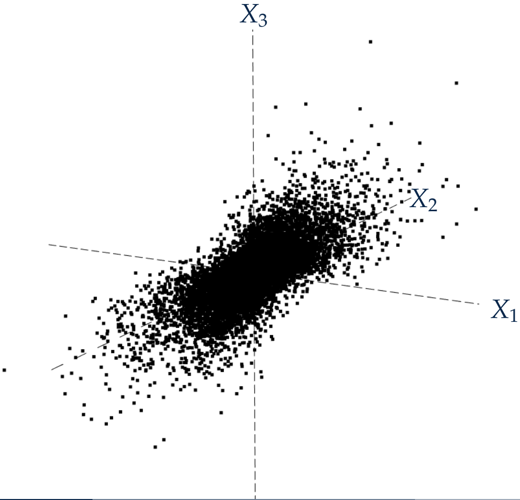
$n = 1000$



Let $\boldsymbol{X}_1, \boldsymbol{X}_2, \ldots \in \mathbb{R}^d$ be iid draws from $\mathbb{P}_{\boldsymbol{X}}$ with standard Laplace marginal distributions, and

$$N_n := \left\{ \frac{\boldsymbol{X}_1}{\log(n)}, \ldots, \frac{\boldsymbol{X}_n}{\log(n)} \right\}$$

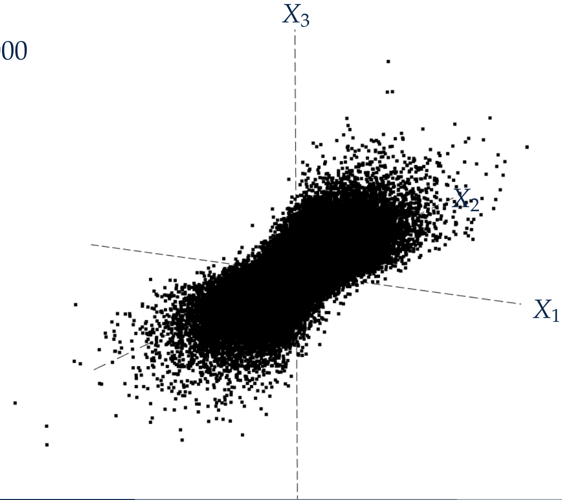
$n = 10000$



Let $\boldsymbol{X}_1, \boldsymbol{X}_2, \ldots \in \mathbb{R}^d$ be iid draws from $\mathbb{P}_{\boldsymbol{X}}$ with standard Laplace marginal distributions, and

$$N_n := \left\{ \frac{\boldsymbol{X}_1}{\log(n)}, \ldots, \frac{\boldsymbol{X}_n}{\log(n)} \right\}$$

$n = 100000$

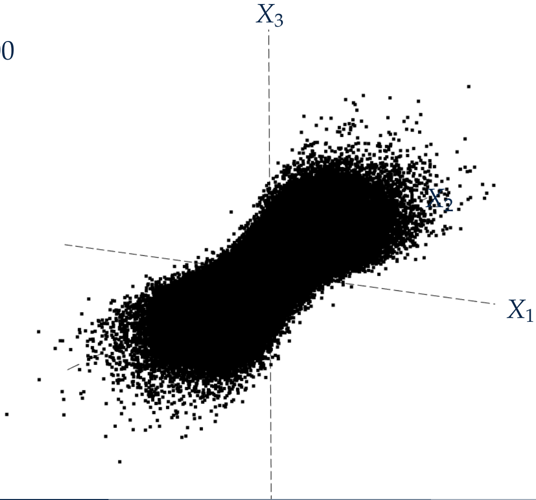


Limit sets

Let $\boldsymbol{X}_1, \boldsymbol{X}_2, \ldots \in \mathbb{R}^d$ be iid draws from $\mathbb{P}_{\boldsymbol{X}}$ with standard Laplace marginal distributions, and

$$N_n := \left\{ \frac{\boldsymbol{X}_1}{\log(n)}, \ldots, \frac{\boldsymbol{X}_n}{\log(n)} \right\}$$

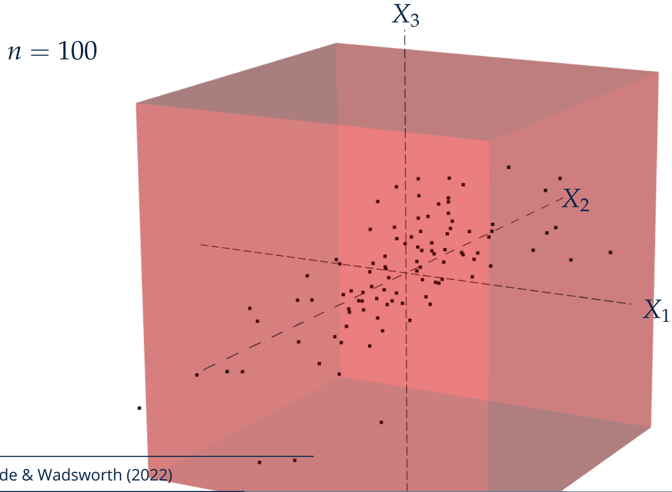
$n = 1000000$



Limit sets

Let $X_1, X_2, \ldots \in \mathbb{R}^d$ be iid draws from $\mathbb{P}_X$ with standard Laplace marginal distributions, and

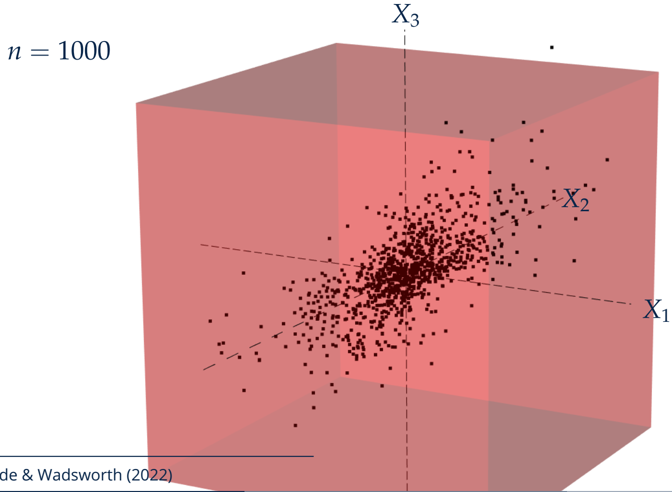
$$N_n := \left\{ \frac{X_1}{\log(n)}, \ldots, \frac{X_n}{\log(n)} \right\} \xrightarrow{\mathbb{P}} \mathcal{G} \subseteq [-1, 1]^d$$



Nolde & Wadsworth (2022)

Let $\boldsymbol{X}_1, \boldsymbol{X}_2, \ldots \in \mathbb{R}^d$ be iid draws from $\mathbb{P}_{\boldsymbol{X}}$ with standard Laplace marginal distributions, and

$$N_n := \left\{ \frac{\boldsymbol{X}_1}{\log(n)}, \ldots, \frac{\boldsymbol{X}_n}{\log(n)} \right\} \xrightarrow{\mathbb{P}} \mathcal{G} \subseteq [-1, 1]^d$$



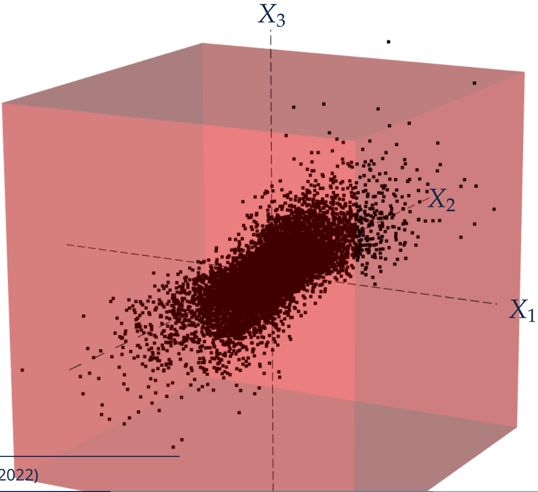
Nolde & Wadsworth (2022)

Limit sets

Let $\boldsymbol{X}_1, \boldsymbol{X}_2, \ldots \in \mathbb{R}^d$ be iid draws from $\mathbb{P}_{\boldsymbol{X}}$ with standard Laplace marginal distributions, and

$$N_n := \left\{ \frac{\boldsymbol{X}_1}{\log(n)}, \ldots, \frac{\boldsymbol{X}_n}{\log(n)} \right\} \xrightarrow{\mathbb{P}} \mathcal{G} \subseteq [-1, 1]^d$$

$n = 10000$



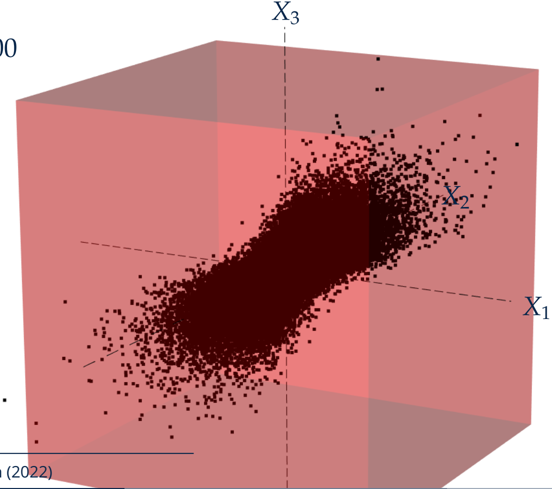
Nolde & Wadsworth (2022)

Limit sets

Let $X_1, X_2, \ldots \in \mathbb{R}^d$ be iid draws from $\mathbb{P}_X$ with standard Laplace marginal distributions, and

$$N_n := \left\{ \frac{X_1}{\log(n)}, \ldots, \frac{X_n}{\log(n)} \right\} \xrightarrow{\mathbb{P}} \mathcal{G} \subseteq [-1, 1]^d$$

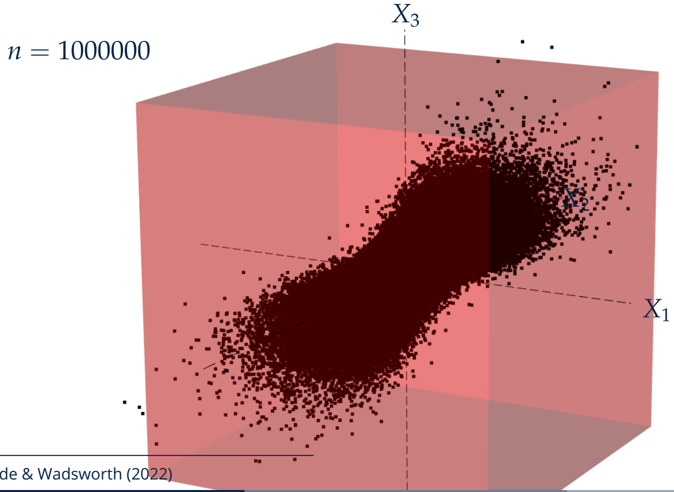
$n = 100000$



Nolde & Wadsworth (2022)

Let $X_1, X_2, \ldots \in \mathbb{R}^d$ be iid draws from $\mathbb{P}_X$ with standard Laplace marginal distributions, and

$$N_n := \left\{ \frac{X_1}{\log(n)}, \ldots, \frac{X_n}{\log(n)} \right\} \xrightarrow{\mathbb{P}} \mathcal{G} \subseteq [-1, 1]^d$$



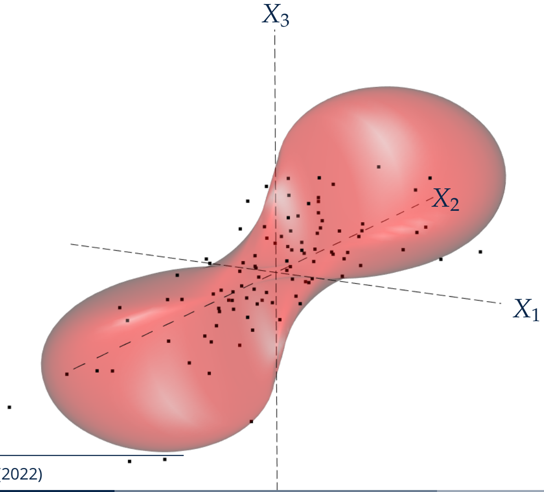
Nolde & Wadsworth (2022)

Limit sets

Let $\boldsymbol{X}_1, \boldsymbol{X}_2, \ldots \in \mathbb{R}^d$ be iid draws from $\mathbb{P}_{\boldsymbol{X}}$ with standard Laplace marginal distributions, and

$$N_n := \left\{ \frac{\boldsymbol{X}_1}{\log(n)}, \ldots, \frac{\boldsymbol{X}_n}{\log(n)} \right\} \xrightarrow{\mathbb{P}} \mathcal{G} \subseteq [-1,1]^d$$

$n = 100$

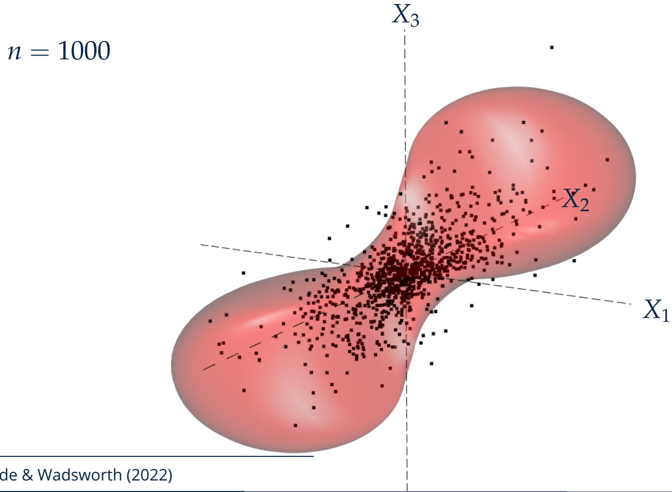


Nolde & Wadsworth (2022)

Limit sets

Let $\boldsymbol{X}_1, \boldsymbol{X}_2, \ldots \in \mathbb{R}^d$ be iid draws from $\mathbb{P}_{\boldsymbol{X}}$ with standard Laplace marginal distributions, and

$$N_n := \left\{ \frac{\boldsymbol{X}_1}{\log(n)}, \ldots, \frac{\boldsymbol{X}_n}{\log(n)} \right\} \xrightarrow{\mathbb{P}} \mathcal{G} \subseteq [-1, 1]^d$$

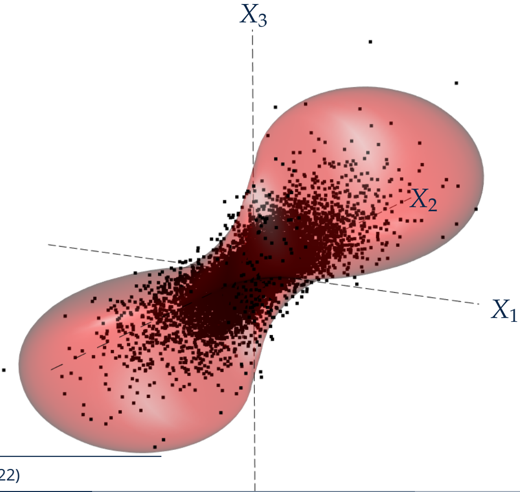


Nolde & Wadsworth (2022)

Let $\boldsymbol{X}_1, \boldsymbol{X}_2, \ldots \in \mathbb{R}^d$ be iid draws from $\mathbb{P}_{\boldsymbol{X}}$ with standard Laplace marginal distributions, and

$$N_n := \left\{ \frac{\boldsymbol{X}_1}{\log(n)}, \ldots, \frac{\boldsymbol{X}_n}{\log(n)} \right\} \xrightarrow{\mathbb{P}} \color{red}{\mathcal{G}} \subseteq [-1, 1]^d$$

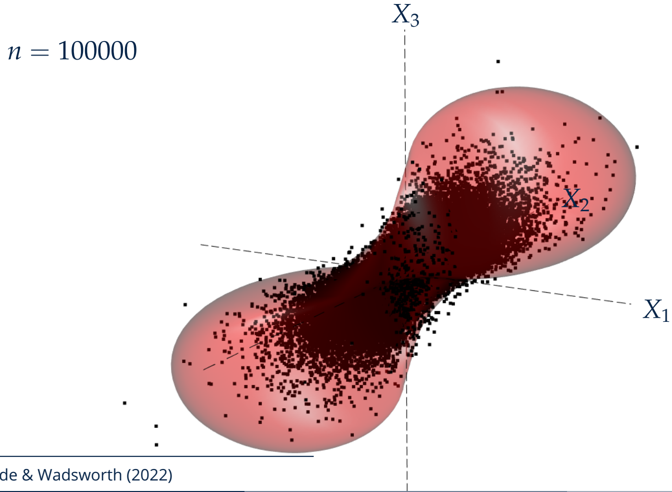
$n = 10000$



Nolde & Wadsworth (2022)

Let $X_1, X_2, \ldots \in \mathbb{R}^d$ be iid draws from $\mathbb{P}_X$ with standard Laplace marginal distributions, and

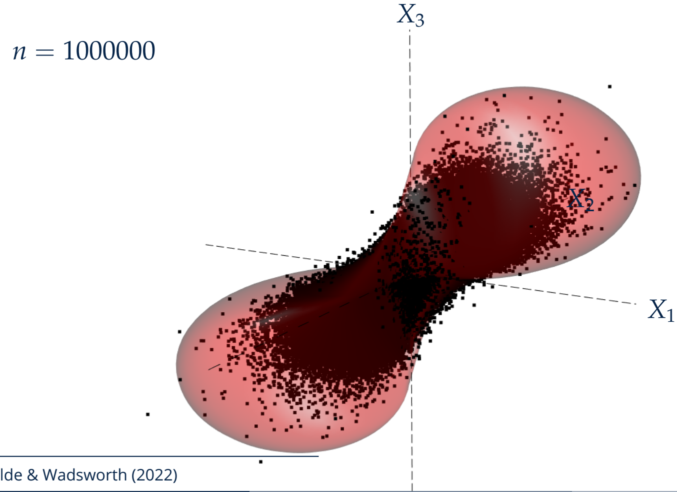
$$N_n := \left\{ \frac{X_1}{\log(n)}, \ldots, \frac{X_n}{\log(n)} \right\} \xrightarrow{\mathbb{P}} \mathcal{G} \subseteq [-1,1]^d$$



Nolde & Wadsworth (2022)

Let $\boldsymbol{X}_1, \boldsymbol{X}_2, \ldots \in \mathbb{R}^d$ be iid draws from $\mathbb{P}_{\boldsymbol{X}}$ with standard Laplace marginal distributions, and

$$N_n := \left\{ \frac{\boldsymbol{X}_1}{\log(n)}, \ldots, \frac{\boldsymbol{X}_n}{\log(n)} \right\} \xrightarrow{\mathbb{P}} \mathcal{G} \subseteq [-1, 1]^d$$



Nolde & Wadsworth (2022)

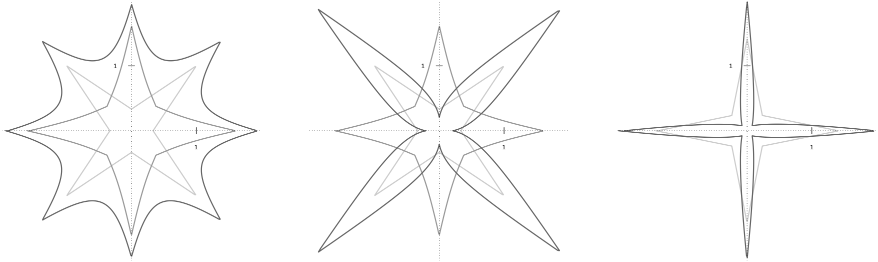
Starshaped sets ★ – A basis for our model construction

- A set $\mathcal{B} \in \mathbb{R}^d$ is starshaped if there exists a set $\ker(\mathcal{B}) \subseteq \mathcal{B}$ such that for $\boldsymbol{x} \in \ker(\mathcal{B})$ and for all $\boldsymbol{y} \in \mathcal{B}$, the segment $[\boldsymbol{x} : \boldsymbol{y}] \in \mathcal{B}$.

- A set $\mathcal{B} \in \bigstar$ is in one-to-one correspondence with a radial function

$$r_{\mathcal{B}}(\boldsymbol{w}) = \sup\{\lambda \in \mathbb{R} : \lambda \boldsymbol{w} \in \mathcal{B}\}, \quad \boldsymbol{w} \in \mathbb{S}^{d-1}.$$

- Starshaped sets admit algebraic operations via their radial functions:



[1] Hansen et al. (2020)

Operations on starshaped sets[1]

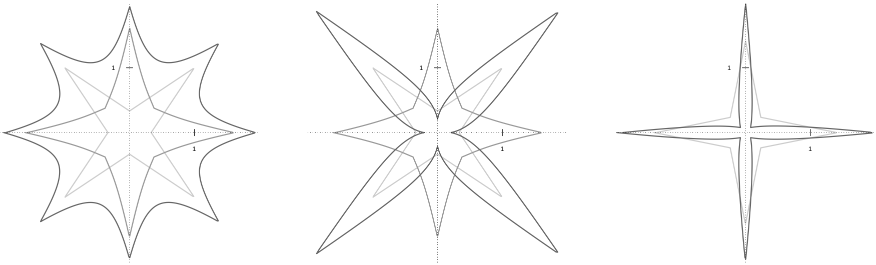
Example

Let $\mathcal{B}_1$ and $\mathcal{B}_2$ be starshaped sets, then

i) $\mathcal{B} = \mathcal{B}_1 + \mathcal{B}_2$ *has radial function* $r_{\mathcal{B}} = r_{\mathcal{B}_1} + r_{\mathcal{B}_2}$.

ii) $\mathcal{B} = \mathcal{B}_1 \cdot \mathcal{B}_2$ *has radial function* $r_{\mathcal{B}} = r_{\mathcal{B}_1} r_{\mathcal{B}_2}$.

iii) $\mathcal{B} = \mathcal{B}_1^d$ *has radial function* $r_{\mathcal{B}} = r_{\mathcal{B}_1}^d$.



[1] Hansen et al. (2020)

Limit sets

A sufficient condition on f_X for N_n to converge onto $\mathcal{G}$ is that

$$-\frac{\log f_X(t\boldsymbol{x}_t)}{t} \to g_{\mathcal{G}}(\boldsymbol{x}), \quad \boldsymbol{x}_t \to \boldsymbol{x}, \text{ as } t \to \infty, \quad \boldsymbol{x} \in \mathbb{R}^d, \tag{1}$$

for a continuous gauge function $g_{\mathcal{G}} : \mathbb{R}^d \to \mathbb{R}_{\geq 0}$. Then, $\mathcal{G} \in \bigstar$ and it has radial function $r_{\mathcal{G}} : \mathbb{S}^{d-1} \to \mathbb{R}_{\geq 0}$ given by $r_{\mathcal{G}} = 1/g_{\mathcal{G}}$.

Our limit set $\mathcal{G}$ can be defined by

$$\mathcal{G} = \{\boldsymbol{x} \in \mathbb{R} : r_{\mathcal{G}}(\boldsymbol{x}) \geq 1\}.$$

Nolde & Wadsworth (2022)

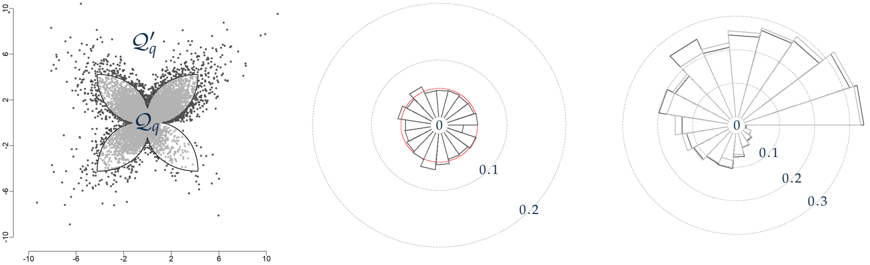
The quantile set $\mathcal{Q}_q$

- We let $\mathcal{Q}_q$ via the q-th quantile of $R \mid \boldsymbol{W} = \boldsymbol{w}$, that is, it satisfies

$$\mathbb{P}[R \leq r_{\mathcal{Q}_q}(\boldsymbol{w}) \mid W = \boldsymbol{w}] = q, \quad \text{for all } \boldsymbol{w} \in \mathbb{S}^{d-1}.$$

- $\mathcal{Q}_q$ then satisfies that

$$\mathbb{P}[\boldsymbol{X} \notin \mathcal{Q}_q] = 1 - q, \qquad \text{and} \qquad \boldsymbol{W} \mid \{\boldsymbol{X} \notin \mathcal{Q}_q\} \stackrel{d}{=} \boldsymbol{W}.$$



Left: Independent samples ($n = 2 \times 10^4$) from a bivariate distribution having true quantile set $\mathcal{Q}_{0.95}$, boundary $\partial\mathcal{Q}_{0.95}$ (solid black line) and complement $\mathcal{Q}'_{0.95}$. Centre: Empirical proportion of exceedances binned by angular regions with true exceedance probability (0.05) in red. Right: Circular histogram of the density of all sampled angles (light grey) and of exceedance angles (dark grey) with concentric circles denoting density level sets.

Exceedances of $\mathcal{Q}_q$

- Note that the event $\{\boldsymbol{X} = R\boldsymbol{W} \notin \mathcal{Q}_q\}$ corresponds to $\{R > r_{\mathcal{Q}_q}(\boldsymbol{W})\}$.
- Papastathopoulos et al. (2023) show conditions under which there exist a starshaped set $\mathcal{G}$ such that

$$\left(\frac{R - r_{\mathcal{Q}_q}(\boldsymbol{W})}{r_{\mathcal{G}}(\boldsymbol{W})}, \boldsymbol{W}\right) \Big| \{R > r_{\mathcal{Q}_q}(\boldsymbol{W})\} \xrightarrow{d} (Z, \boldsymbol{V}), \quad \text{as } q \to 1, \tag{2}$$

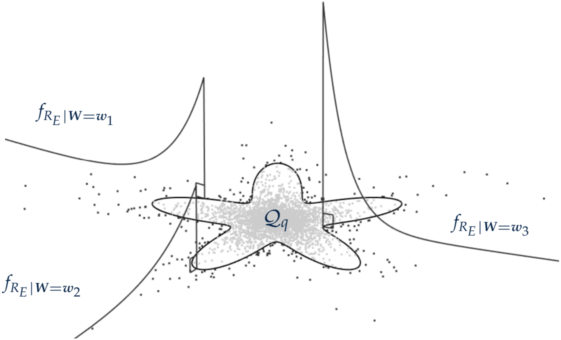
where $Z \sim \text{Exp}(1)$ and $\boldsymbol{V} \sim \mathbb{P}_{\boldsymbol{W}}$.

Exceedances of $\mathcal{Q}_q$

- Note that the event $\{\boldsymbol{X} = R\boldsymbol{W} \notin \mathcal{Q}_q\}$ corresponds to $\{R > r_{\mathcal{Q}_q}(\boldsymbol{W})\}$.
- Papastathopoulos et al. (2023) show conditions under which there exist a starshaped set $\mathcal{G}$ such that

$$\left(\frac{R - r_{\mathcal{Q}_q}(\boldsymbol{W})}{r_{\mathcal{G}}(\boldsymbol{W})}, \boldsymbol{W}\right) \mid \{R > r_{\mathcal{Q}_q}(\boldsymbol{W})\} \xrightarrow{d} (Z, \boldsymbol{V}), \quad \text{as } q \to 1, \tag{3}$$

where $Z \sim \text{Exp}(1)$ and $\boldsymbol{V} \sim \mathbb{P}_{\boldsymbol{W}}$.



PROPOSED MODELS

Imposing structure on $\mathcal{Q}_q$, $\mathcal{G}$, and $\mathcal{W}$

Links between parameters and models

Under appropriate convergence conditions[1], it can be shown that the quantile set $\mathcal{Q}_q$ is asymptotically a scale multiple of the scaling/limit set $\mathcal{G}$, that is,

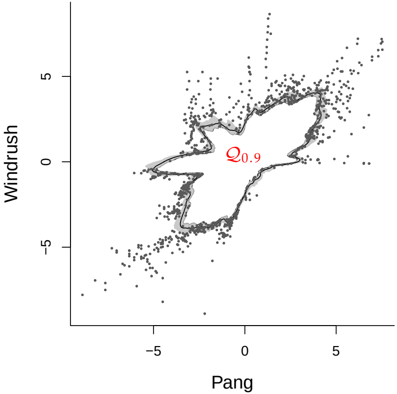
$$\mathcal{Q}_q \approx \alpha_q \mathcal{G}, \quad \alpha_q > 0, \quad \text{as } q \to 1$$

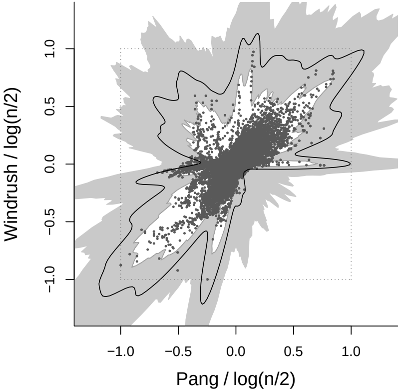
[1] Wadsworth & Campbell (2024)

Links between parameters and models

Under appropriate convergence conditions[1], it can be shown that the quantile set $\mathcal{Q}_q$ is asymptotically a scale multiple of the scaling/limit set $\mathcal{G}$, that is,

$$\mathcal{Q}_q \approx \alpha_q \mathcal{G}, \quad \alpha_q > 0, \quad \text{as } q \to 1$$





[1] Wadsworth & Campbell (2024)

Links between parameters and models

If the density of $\boldsymbol{X}$ is homothetic with respect to $r_{\mathcal{G}}^{-1}$, that is,

$$f_{\boldsymbol{X}}(\boldsymbol{x}) = h_0(r_{\mathcal{G}}^{-1}(\boldsymbol{x})), \quad x \in \mathbb{R}^d,$$

for a positive, decreasing, and continuous function h_0, then $\mathcal{G}$ and $\mathcal{W}$ can be linked[1] through

$$r_{\mathcal{W}}(\boldsymbol{w}) = f_{\boldsymbol{W}}(\boldsymbol{w}) = \frac{r_{\mathcal{G}}(\boldsymbol{w})^d}{d|\mathcal{G}|}, \quad \boldsymbol{w} \in \mathbb{S}^{d-1}.$$

[1] Papastathopoulos et al. (2023)

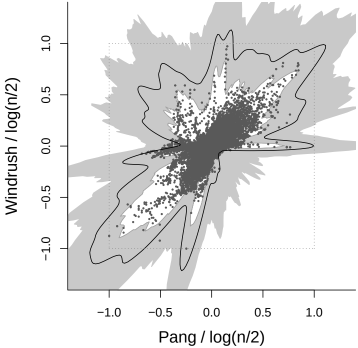
Links between parameters and models

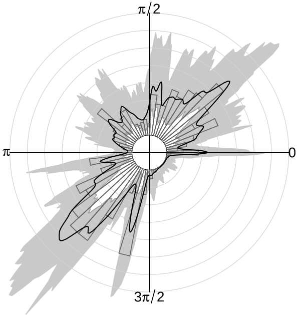
If the density of $\boldsymbol{X}$ is homothetic with respect to $r_{\mathcal{G}}^{-1}$, that is,

$$f_{\boldsymbol{X}}(\boldsymbol{x}) = h_0(r_{\mathcal{G}}^{-1}(\boldsymbol{x})), \quad x \in \mathbb{R}^d,$$

for a positive, decreasing, and continuous function h_0, then $\mathcal{G}$ and $\mathcal{W}$ can be linked[1] through

$$r_{\mathcal{W}}(\boldsymbol{w}) = f_{\boldsymbol{W}}(\boldsymbol{w}) = \frac{r_{\mathcal{G}}(\boldsymbol{w})^d}{d|\mathcal{G}|}, \quad \boldsymbol{w} \in \mathbb{S}^{d-1}.$$





[1] Papastathopoulos et al. (2023)

Imposing structure on $\mathcal{Q}_q$, $\mathcal{G}$, and $\mathcal{W}$

- Any positive function $r_{\mathcal{B}}$ defined on $\mathbb{S}^{d-1}$ can be written as

$$r_{\mathcal{B}}(\boldsymbol{w}) = \beta_{\mathcal{B}} f_{\mathcal{B}}(\boldsymbol{w}), \quad \boldsymbol{w} \in \mathbb{S}^{d-1},$$

  for a constant $\beta_{\mathcal{B}} = \int_{\mathbb{S}^{d-1}} r_{\mathcal{B}}(\boldsymbol{w})\,\mathrm{d}\boldsymbol{w}$ and density $f_{\mathcal{B}}$ integrating to 1 on $\mathbb{S}^{d-1}$.

- Using the links $\mathcal{G}$–$\mathcal{Q}_q$ and $\mathcal{G}$–$\mathcal{W}$, we can formulate a statistical model

$$r_{\mathcal{Q}_q}(\boldsymbol{w}) = \beta_{\mathcal{Q}_q} f_{\mathbf{W}}(\boldsymbol{w})^d \text{ and } r_{\mathcal{G}}(\boldsymbol{w}) = \beta_{\mathcal{G}} f_{\mathbf{W}}(\boldsymbol{w})^d, \qquad \boldsymbol{w} \in \mathbb{S}^{d-1}.$$

Imposing structure on $\mathcal{Q}_q$, $\mathcal{G}$, and $\mathcal{W}$

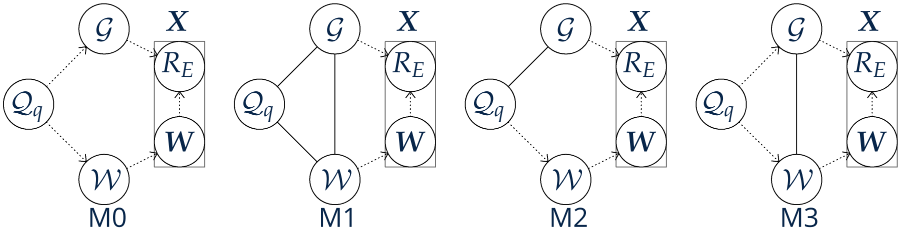
- Any positive function $r_{\mathcal{B}}$ defined on $\mathbb{S}^{d-1}$ can be written as

$$r_{\mathcal{B}}(\boldsymbol{w}) = \beta_{\mathcal{B}} f_{\mathcal{B}}(\boldsymbol{w}), \quad \boldsymbol{w} \in \mathbb{S}^{d-1},$$

for a constant $\beta_{\mathcal{B}} = \int_{\mathbb{S}^{d-1}} r_{\mathcal{B}}(\boldsymbol{w})\, \mathrm{d}\boldsymbol{w}$ and density $f_{\mathcal{B}}$ integrating to 1 on $\mathbb{S}^{d-1}$.

- Using the links $\mathcal{G}$–$\mathcal{Q}_q$ and $\mathcal{G}$–$\mathcal{W}$, we can formulate a statistical model

$$r_{\mathcal{Q}_q}(\boldsymbol{w}) = \beta_{\mathcal{Q}_q} f_{\boldsymbol{W}}(\boldsymbol{w})^d \text{ and } r_{\mathcal{G}}(\boldsymbol{w}) = \beta_{\mathcal{G}} f_{\boldsymbol{W}}(\boldsymbol{w})^d, \qquad \boldsymbol{w} \in \mathbb{S}^{d-1}.$$



STATISTICAL INFERENCE

Normalising flows[1] and density estimation[2]

- A normalising flow (NF) learns a transformation mapping a random variable $\boldsymbol{Y} \in \mathcal{Y}$ with unknown distribution to that of a known, base variable $\boldsymbol{Z} \in \mathcal{Z}$.

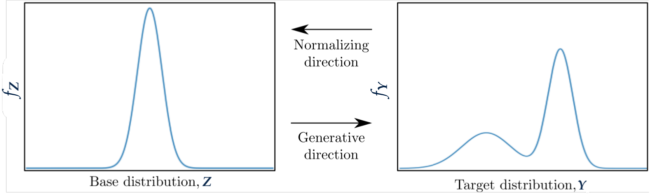


Figure 1 of Kobyzev et al. (2021)

[1] Tabak & Vanden-Eijnden (2010), [2] Dinh et al. (2015)

Normalising flows[1] and density estimation[2]

- A normalising flow (NF) learns a transformation mapping a random variable $\boldsymbol{Y} \in \mathcal{Y}$ with unknown distribution to that of a known, base variable $\boldsymbol{Z} \in \mathcal{Z}$.

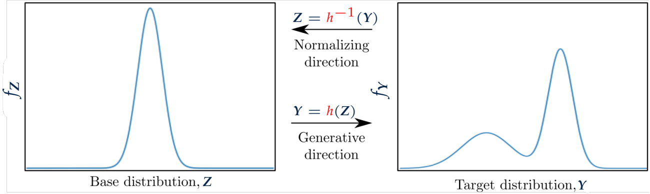


Figure 1 of Kobyzev et al. (2021)

- Assuming Y admits a density on $\mathcal{Y}$, this problem can be phrased as aiming to infer a (bijective and differentiable) transformation function h such that

$$f_{\boldsymbol{Y}}(\boldsymbol{y}) = f_{\boldsymbol{Z}}\{h^{-1}(\boldsymbol{y})\} \left| \frac{\partial h^{-1}(\boldsymbol{y})}{\partial \boldsymbol{y}} \right|, \quad \boldsymbol{y} \in \mathcal{Y}.$$

In practice, h is modelled as a composition of many simple bijective transformations $h_1, \dots, h_k$, *i.e.* $h = h_1 \circ h_2 \circ \dots \circ h_k$.

[1]Tabak & Vanden-Eijnden (2010), [2]Dinh et al. (2017)

A map from the hypersphere to the hypercylinder

- Transform the observations and models from $\mathbb{S}^{d-1}$ to a cylindrical space $\mathbb{C}^{d-1}$ (by abuse of notation) via a map T

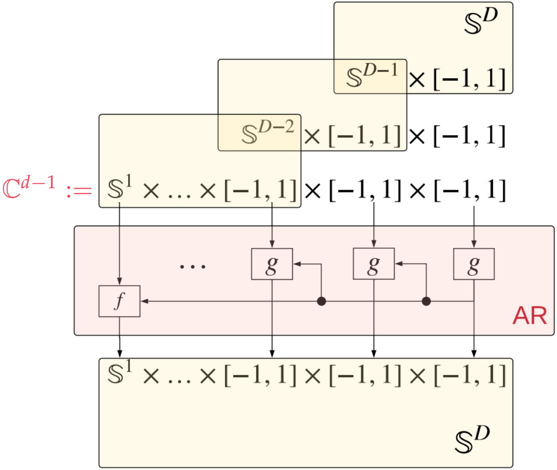
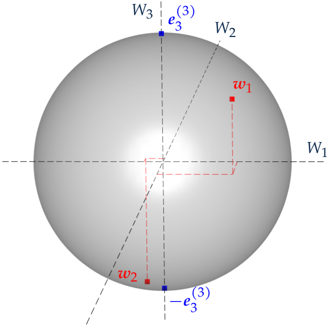


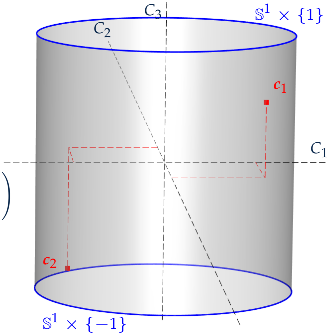
Figure 6 of Rezende et al. (2020)

A map from the hypersphere $\mathbb{S}^2$ to the hypercylinder $\mathbb{C}^2$



$$T : \mathbb{S}^2 \backslash \mathcal{S}_3 \to \mathbb{C}^2$$

$$\boldsymbol{w} \mapsto \boldsymbol{c} = \left(\frac{w_1}{\sqrt{1-w_3^2}}, \frac{w_2}{\sqrt{1-w_3^2}}, w_3 \right)$$



A model for PDFs and positive functions on $\mathbb{S}^{d-1}$

- It follows from the map T that a target PDF $f_{\mathcal{B}} : \mathbb{S}^{d-1} \setminus \mathcal{S}_d \to [0, \infty)$, describing the shape of a starshaped set $\mathcal{B} \in \mathbb{R}^d$ *a.e.*, can be written as

$$f_{\mathcal{B}}(\boldsymbol{w}) = f_{\boldsymbol{Y}}(T(\boldsymbol{w}))|\partial T(\boldsymbol{w})/\partial \boldsymbol{w}|, \quad \boldsymbol{w} \in \mathbb{S}^{d-1} \setminus \mathcal{S}_d,$$

  for a target PDF $f_{\boldsymbol{Y}}$ defined on $\mathbb{C}^{d-1}$.

- Using the NFs formulation, $f_{\mathcal{B}}$ can in turn be modelled in terms of a known base PDF $f_{\boldsymbol{Z}} : \mathbb{C}^{d-1} \to [0, \infty)$ and a normalising flow $h_{\mathcal{B}}$ as

$$f_{\mathcal{B}}(\boldsymbol{w}) = f_{\boldsymbol{Z}}\{h_{\mathcal{B}}^{-1}(T(\boldsymbol{w}))\} \left| \frac{\partial h_{\mathcal{B}}^{-1}(T(\boldsymbol{w}))}{\partial T(\boldsymbol{w})} \right| \left| \frac{\partial T(\boldsymbol{w})}{\partial \boldsymbol{w}} \right|, \quad \boldsymbol{w} \in \mathbb{S}^{d-1} \setminus \mathcal{S}_d,$$

  where $|\partial T(\boldsymbol{w})/\boldsymbol{w}|$ is the Jacobian of the recursive transformation T.

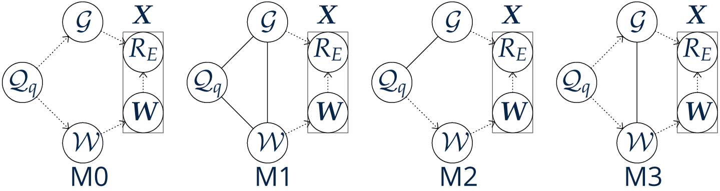
- Further, a model for any positive/radial function $r_{\mathcal{B}}$ of a starshaped set $\mathcal{B}$ – such as the quantile set $\mathcal{Q}_q$ or the scaling set $\mathcal{G}$ – can be obtained via

$$r_{\mathcal{B}} = \beta_{\mathcal{B}} f_{\mathcal{B}}$$

  where $f_{\mathcal{B}}$ is as above, and $\beta_{\mathcal{B}} > 0$ is a coefficient to be learned alongside the NF $h_{\mathcal{B}}$.

Model fitting via loss minimisation

- Recall models M0 to M3:



- Model M0 is fitted by sequentially minimising the losses
  - for $\mathcal{Q}_q$:

$$\mathcal{L}_{\mathcal{Q}_q}(\beta_{\mathcal{Q}_q}, f_{\mathcal{Q}_q}\,;\,\underline{\boldsymbol{x}}) = \frac{1}{n}\sum_{i=1}^{n} \max\left\{(1-q)\left[\|\boldsymbol{x}_i\| - \beta_{\mathcal{Q}_q} f_{\mathcal{Q}_q}\left(\frac{\boldsymbol{x}_i}{\|\boldsymbol{x}_i\|}\right)\right], q\left[\|\boldsymbol{x}_i\| - \beta_{\mathcal{Q}_q} f_{\mathcal{Q}_q}\left(\frac{\boldsymbol{x}_i}{\|\boldsymbol{x}_i\|}\right)\right]\right\}.$$

  - for $\mathcal{G}$:

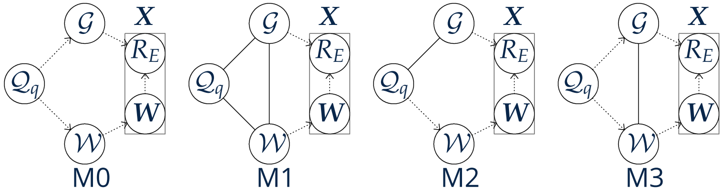
$$\mathcal{L}_{\mathcal{G}}(\beta_{\mathcal{G}}, f_{\mathcal{G}}\,;\, r_{\hat{\mathcal{Q}}_q}, \underline{\boldsymbol{x}}) = -\frac{1}{\#\mathcal{E}}\sum_{i\in\mathcal{E}} \log\left[\{\beta_{\mathcal{G}} f_{\mathcal{G}}(\boldsymbol{x}_i/\|\boldsymbol{x}_i\|)\}^{-1} \exp\left\{-\frac{\|\boldsymbol{x}_i\| - r_{\hat{\mathcal{Q}}_q}(\boldsymbol{x}_i/\|\boldsymbol{x}_i\|)}{\beta_{\mathcal{G}} f_{\mathcal{G}}(\boldsymbol{x}_i/\|\boldsymbol{x}_i\|)}\right\}\right].$$

  - for $\mathcal{W}$:

$$\mathcal{L}_{\mathcal{W}}(f_W\,;\, r_{\hat{\mathcal{Q}}_q}, \underline{\boldsymbol{x}}) = -\frac{1}{\#\mathcal{E}}\sum_{i\in\mathcal{E}} \log f_W(\boldsymbol{x}_i/\|\boldsymbol{x}_i\|).$$

Model fitting via loss minimisation

- Recall models M0 to M3:



- Model M1 is fitted by sequentially minimising the loss $\mathcal{L}_{\mathcal{Q}_q,\mathcal{G},\mathcal{W}}(\beta_{\mathcal{Q}_q},\beta_{\mathcal{G}},f_{\boldsymbol{W}}\,;\,\underline{\boldsymbol{x}}) =$

$$= \mathcal{L}_{\mathcal{Q}_q}(\beta_{\mathcal{Q}_q}, f_{\boldsymbol{W}}^{1/d}\,;\,\underline{\boldsymbol{x}}) + \lambda\big[\mathcal{L}_{\mathcal{G}}(\beta_{\mathcal{G}}, f_{\boldsymbol{W}}^{1/d}\,;\,\beta_{\mathcal{Q}_q} f_{\boldsymbol{W}}^{1/d}, \underline{\boldsymbol{x}}) + \mathcal{L}_{\mathcal{W}}(f_{\boldsymbol{W}}\,;\,\beta_{\mathcal{Q}_q} f_{\boldsymbol{W}}^{1/d}, \underline{\boldsymbol{x}})\big].$$

- The model is wholly defined in terms of only one density $f_{\boldsymbol{W}}$ and two scalars $\beta_{\mathcal{Q}_q}$ and $\beta_{\mathcal{G}}$.

- λ is a weighting hyperparameter accounting for the different scales of the values of the losses.

- Comments on M2 and M3.

A GRADIENT DESCENT APPROACH

A PyTorch[1] implementation[2] of NFs and composite loss minimisation via the Adam optimiser[3]

Data are mollified[4] during training. At the jth of J gradient descent epoch, we use the mollified dataset

$$\underline{\boldsymbol{x}}_{\mathrm{T},j} = \left\{ \|\boldsymbol{x}_i\| \boldsymbol{w}_{i,\varepsilon} \ : \ \boldsymbol{w}_{i,\varepsilon} \sim \mathrm{vonMises}(\boldsymbol{x}_i / \|\boldsymbol{x}_i\|, \sigma_j),\, \boldsymbol{x}_i \in \underline{\boldsymbol{x}}_{\mathrm{T}} \right\}, \tag{4}$$

where vonMises($\boldsymbol{\mu}$,σ) denotes the von Mises distribution with location $\boldsymbol{\mu} \in \mathbb{S}^{d-1}$ and dispersion $\sigma \in \mathbb{R}_{>0}$.

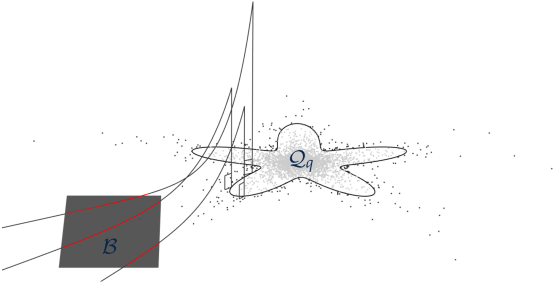
[1] Paszke et al. (2019), [2] Stimper et al. (2023), [3] Kingma & Ba (2017), [4] Tran et al. (2023)

PROBABILITY ESTIMATION

Probability estimation

- For any Borel set $\mathcal{B} \in \mathbb{R}^d \backslash \mathcal{Q}_q$,

$$\mathbb{P}[\boldsymbol{X} \in \mathcal{B} \mid \boldsymbol{X} \notin \mathcal{Q}_q] = \int_{\mathbb{S}^{d-1}} \int_{\mathcal{B} \cap]\mathbf{0}:\boldsymbol{w})} \frac{1}{r_{\mathcal{G}}(\boldsymbol{w})} \exp\left\{ -\frac{r - r_{\mathcal{Q}_q}(\boldsymbol{w})}{r_{\mathcal{G}}(\boldsymbol{w})} \right\} f_{\boldsymbol{W}}(\boldsymbol{w}) \mathrm{d}r \, \mathrm{d}\boldsymbol{w}.$$

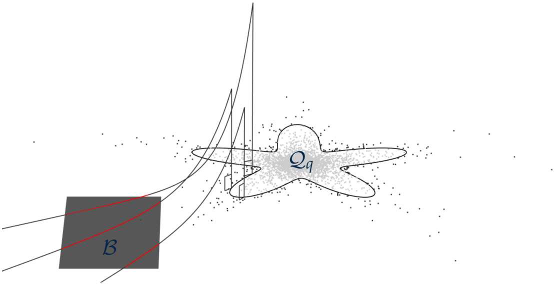


Probability estimation

- For any Borel set $\mathcal{B} \in \mathbb{R}^d \backslash \mathcal{Q}_q$, we use the Monte Carlo integration

$$\mathbb{P}[\boldsymbol{X} \in \mathcal{B} \mid \boldsymbol{X} \notin \mathcal{Q}_q] \xleftarrow{\mathbb{P}} \frac{1}{m} \sum_{i=1}^{m} \int_{\mathcal{B} \cap]\mathbf{0}:\boldsymbol{w}_i)} \frac{1}{r_{\mathcal{G}}(\boldsymbol{w}_i)} \exp\left\{ -\frac{r - r_{\mathcal{Q}_q}(\boldsymbol{w}_i)}{r_{\mathcal{G}}(\boldsymbol{w}_i)} \right\} \mathrm{d}r, \quad n \to \infty.$$

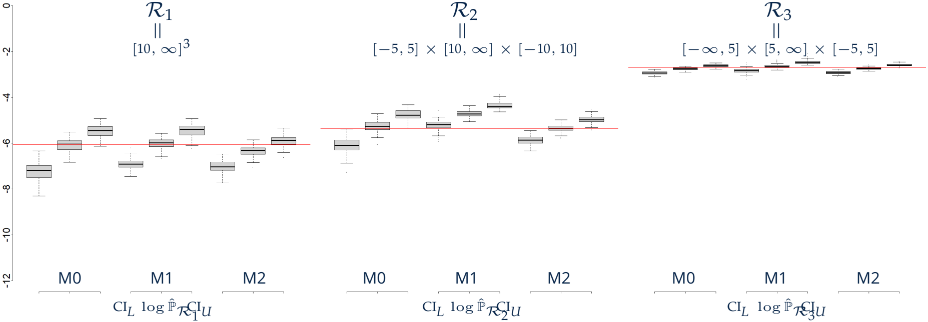
where $\boldsymbol{w}_1, \ldots, \boldsymbol{w}_m \sim f_{\boldsymbol{W}}$.



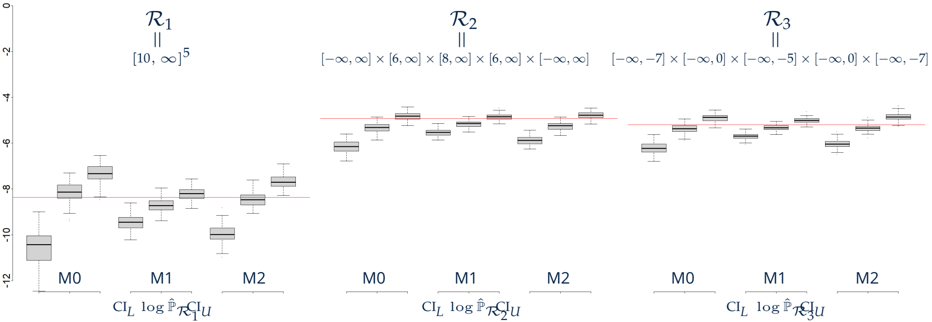
- The integral is exact provided one knows all radial entry and exit points of $\mathcal{B}$.

- The collection $\boldsymbol{w}_1, \ldots, \boldsymbol{w}_m \sim f_{\boldsymbol{W}}$ is sampled fast using the generative direction of the NF.

Simulation study results – 3 dimensions

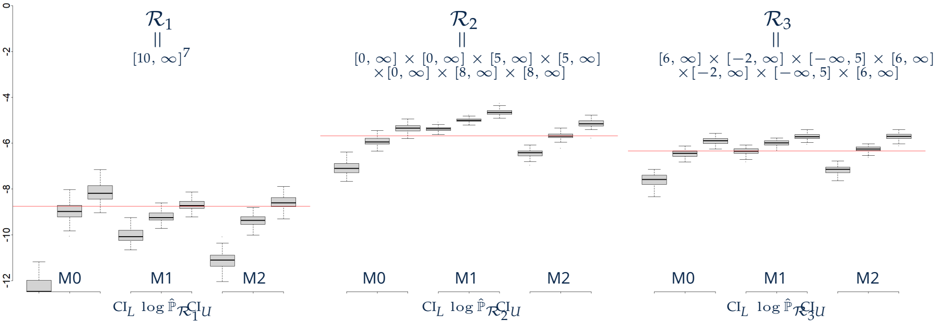


Boxplots of 100 estimated log-probabilities and associated lower- and upper-bounds of 95% bootstrap confidence intervals for the sets $\mathcal{R}_1, \mathcal{R}_2, \mathcal{R}_3 \in \mathbb{R}^3$. ($n = 10^4$).



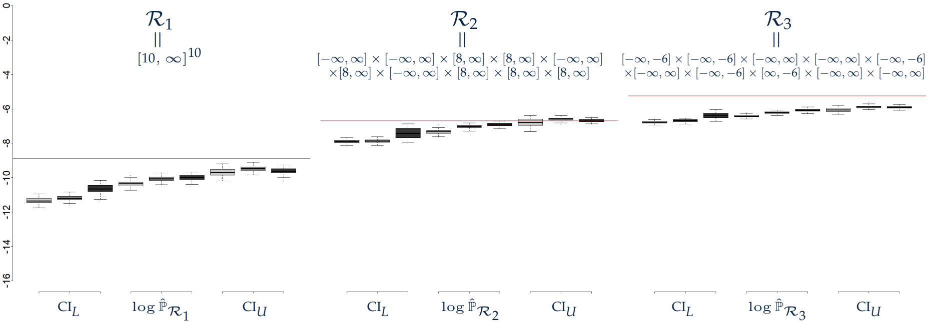
Boxplots of 100 estimated log-probabilities and associated lower- and upper-bounds of 95% bootstrap confidence intervals for the sets $\mathcal{R}_1, \mathcal{R}_2, \mathcal{R}_3 \in \mathbb{R}^5$. ($n = 10^4$).

Simulation study results – 7 dimensions



Boxplots of 100 estimated log-probabilities and associated lower- and upper-bounds of 95% bootstrap confidence intervals for the sets $\mathcal{R}_1, \mathcal{R}_2, \mathcal{R}_3 \in \mathbb{R}^7$. ($n = 10^4$).

Simulation study results – 10 dimensions

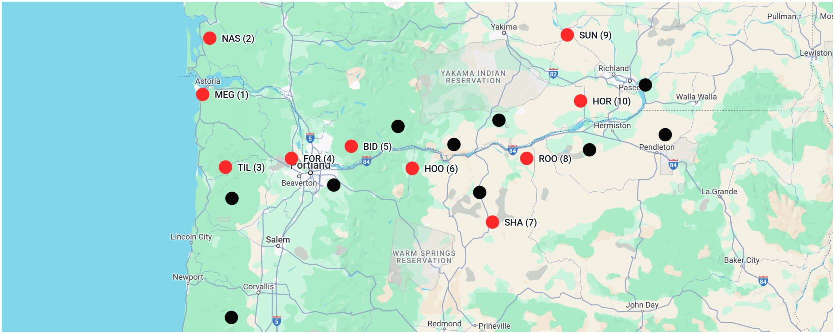


Boxplots of 100 estimated log-probabilities and associated lower- and upper-bounds of 95% bootstrap confidence intervals for the sets $\mathcal{R}_1, \mathcal{R}_2, \mathcal{R}_3 \in \mathbb{R}^{10}$. ($n = 5 \times 10^4$, 10^5, 2×10^5).

LOW AND HIGH WIND SPEEDS

In relation to electricity production in the Pacific Northwest, United States

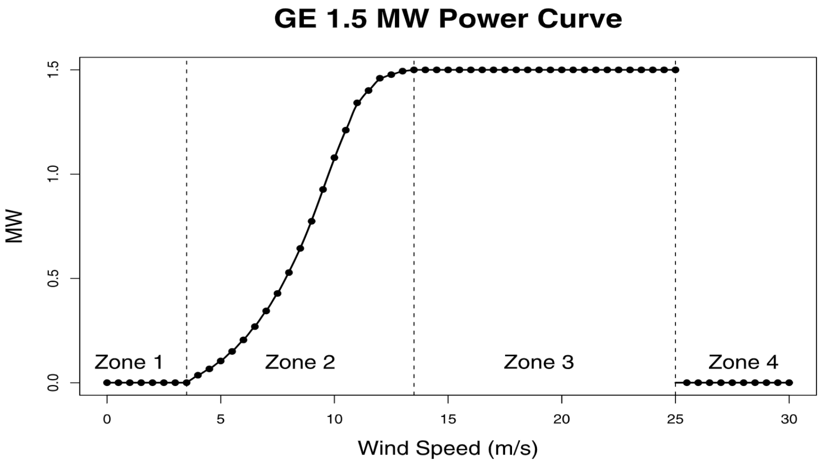
Pacific Northwest region of the United States[1,2]



Hourly windspeeds, Jan 1, 2012–Jan 1, 2015.

[1]Huser et al. (2017), [2]Castro-Camilo et al. (2019)

Wind turbine power curves[1]



[1] Hering & Genton (2010)

Scale-shape homogenisation

- Define the windspeed

$$X^{\mathrm{o}}_{j,m,h}$$

  at station j in month m of the year and hour h.

Scale-shape homogenisation

- Define the windspeed

$$X^{\mathrm{o}}_{j,m,h}$$

at station j in month m of the year and hour h.

- We assume[1]

$$X^{\mathrm{o}}_{j,m,h} \sim \text{Weibull}\big(\lambda_{j,m,h} = s_{j,1}(m) + s_{j,2}(h)\,,\ \kappa_{j,m,h} = s_{j,3}(m) + s_{j,4}(h)\big), \tag{5}$$

where s denotes a cubic cyclic spline on $m \in \{1, \ldots, 12\}$ or $h \in \{0, \ldots, 23\}$.

[1] Elliott et al. (2004)

Scale-shape homogenisation

- Define the windspeed

$$X^{\mathrm{o}}_{j,m,h}$$

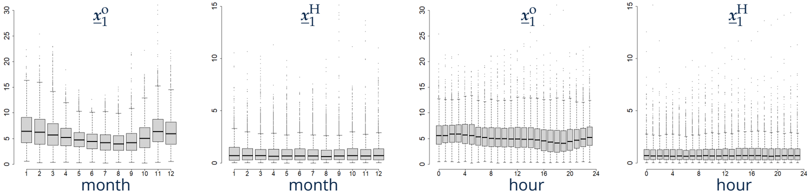
at station j in month m of the year and hour h.

- We assume[1]

$$X^{\mathrm{o}}_{j,m,h} \sim \text{Weibull}\big(\lambda_{j,m,h} = s_{j,1}(m) + s_{j,2}(h)\,,\ \kappa_{j,m,h} = s_{j,3}(m) + s_{j,4}(h)\big), \tag{6}$$

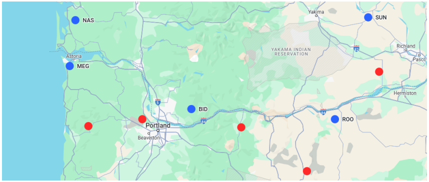
where s denotes a cubic cyclic spline on $m \in \{1, \ldots, 12\}$ or $h \in \{0, \ldots, 23\}$.

- We fit the model using `evgam`[2] and apply $X^{\mathrm{H}}_{j,m,h} := (X^{\mathrm{o}}_{j,m,h}/\hat{\lambda}_{j,m,h})^{\hat{\kappa}_{j,m,h}}$

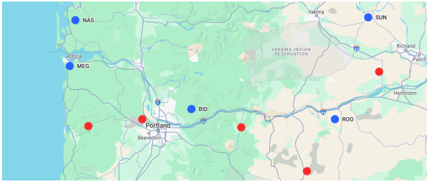


[1] Elliott et al. (2004), [2] Youngman (2022)

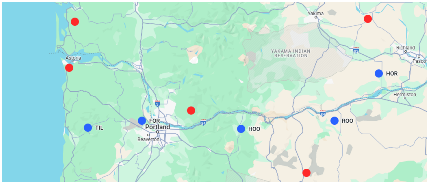
Analysis of station configurations – January at 18:00



(a) Minimises probability of no production

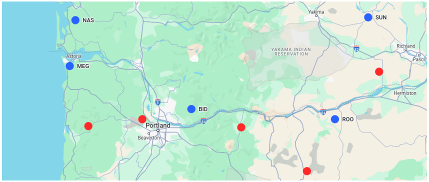


(a) Minimises probability of no production

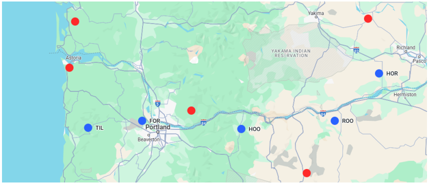


(b) Maximises probability of no production

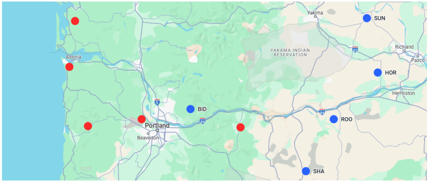
Analysis of station configurations – January at 18:00



(a) Minimises probability of no production

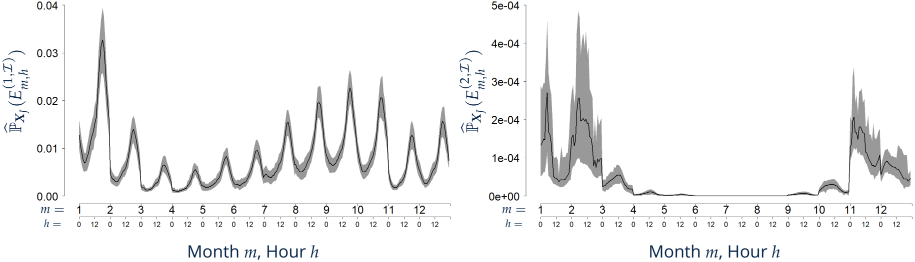


(b) Maximises probability of no production



(c) Maximises probability of full production

Analysis of seasonality of power production – configuration (a)



Configuration (a): Minimises probability of no production

Concluding remarks

The proposed methodology provides

- flexible statistical inference for “high” dimensional random vectors;
- models bridging parsimony and flexibility by exploiting the geometry of 3 structural parameters to improve statistical inference.

Concluding remarks

The proposed methodology provides

- flexible statistical inference for “high” dimensional random vectors;
- models bridging parsimony and flexibility by exploiting the geometry of 3 structural parameters to improve statistical inference.

The framework enables

- fast inference and relatively fast bootstrapping (with possibility of pre-training);
- fast probability estimation enabled by very fast sampling from normalising flows.

See De Monte, L., Huser, R., Papastathopoulos, I., Richards, J. (2025). Generative modelling of multivariate geometric extremes using normalising flows. arXiv:2505.02957.

References

Castro-Camilo, D., Huser, R. & Rue, H. (2019), 'A spliced gamma-generalized Pareto model for short-term extreme wind speed probabilistic forecasting', *J. Agric. Biol. Environ. Stat.* **24**(3), 517–534.

De Monte, L., Huser, R., Papastathopoulos, I. & Richards, J. (2025), 'Generative modelling of multivariate geometric extremes using normalising flows'.

Dinh, L., Krueger, D. & Bengio, Y. (2015), NICE: non-linear independent components estimation, *in* Y. Bengio & Y. LeCun, eds, '3rd International Conference on Learning Representations, ICLR 2015, San Diego, CA, USA, May 7-9, 2015, Workshop Track Proceedings'.

Dinh, L., Sohl-Dickstein, J. & Bengio, S. (2017), 'Density estimation using real nvp'.

Elliott, D., Schwartz, M. & Scott, G. (2004), Wind resource base, *in* C. J. Cleveland, ed., 'Encyclopedia of Energy', Elsevier, New York, pp. 465–479.

Hansen, G., Herburt, I., Martini, H. & Moszyńska, M. (2020), 'Starshaped sets', *Aequationes mathematicae* **94**, 1001–1092.

Hering, A. S. & Genton, M. G. (2010), 'Powering up with space-time wind forecasting', *Journal of the American Statistical Association* **105**(489), 92–104.

Huser, R., Opitz, T. & Thibaud, E. (2017), 'Bridging asymptotic independence and dependence in spatial extremes using gaussian scale mixtures', *Spatial Statistics* **21**, 166–186.

Kingma, D. P. & Ba, J. (2017), 'Adam: A method for stochastic optimization'.

Kobyzev, I., Prince, S. J. & Brubaker, M. A. (2021), 'Normalizing flows: An introduction and review of current methods', *IEEE Transactions on Pattern Analysis and Machine Intelligence* **43**(11), 3964–3979.

Nolde, N. & Wadsworth, J. L. (2022), 'Linking representations for multivariate extremes via a limit set', *Advances in Applied Probability* **54**(3), 688–717.

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

Paszke, A., Gross, S., Massa, F., Lerer, A., Bradbury, J., Chanan, G., Killeen, T., Lin, Z., Gimelshein, N., Antiga, L., Desmaison, A., Köpf, A., Yang, E., DeVito, Z., Raison, M., Tejani, A., Chilamkurthy, S., Steiner, B., Fang, L., Bai, J. & Chintala, S. (2019), 'Pytorch: An imperative style, high-performance deep learning library'.

Rezende, D. J., Papamakarios, G., Racanière, S., Albergo, M. S., Kanwar, G., Shanahan, P. E. & Cranmer, K. (2020), 'Normalizing flows on tori and spheres'.

Stimper, V., Liu, D., Campbell, A., Berenz, V., Ryll, L., Schölkopf, B. & Hernández-Lobato, J. M. (2023), 'normflows: A pytorch package for normalizing flows', *Journal of Open Source Software* **8**(86), 5361.

Tabak, E. G. & Vanden-Eijnden, E. (2010), 'Density estimation by dual ascent of the log-likelihood', *Communications in Mathematical Sciences* **8**(1), 217–233.

Tran, B.-H., Franzese, G., Michiardi, P. & Filippone, M. (2023), One-line-of-code data mollification improves optimization of likelihood-based generative models, *in* A. Oh, T. Naumann, A. Globerson, K. Saenko, M. Hardt & S. Levine, eds, 'Advances in Neural Information Processing Systems', Vol. 36, Curran Associates, Inc., pp. 6545–6567.

Wadsworth, J. L. & Campbell, R. (2024), 'Statistical inference for multivariate extremes via a geometric approach', *Journal of the Royal Statistical Society Series B: Statistical Methodology* p. qkae030.

Youngman, B. D. (2022), 'evgam: An R package for generalized additive extreme value models', *J. Stat. Softw.* **103**(3), 1–26.

Uniform-on-$\mathbb{S}^{d-1}$ penalisation

- To devise the uniform density on $\mathbb{S}^{d-1}$, we consider $A_{d-1}(r)$ the hypervolume (or surface area) of the $(d-1)$-sphere of radius r given by

$$A_{d-1}(r) = \frac{2\pi^{d/2}}{\Gamma(d/2)} r^{d-1}, \quad r \in (0, \infty),$$

  where Γ denotes the gamma function.

- It follows that a PDF with uniform density on $\mathbb{S}^{d-1}$ is given by

$$f_U(\boldsymbol{w}) = 1/A_{d-1}(1)$$

  for all $\boldsymbol{w} \in \mathbb{S}^{d-1}$.

- Penalisation of $f_{\mathcal{D}}$ away from f_U can then be performed via the Kullback–Leibler divergence $\mathrm{D}_{\mathrm{KL}}[f_U \| f_{\mathcal{D}}] = \int_{\mathbb{S}^{d-1}} \log[f_U(\boldsymbol{w})/f_{\mathcal{D}}(\boldsymbol{w})] f_U(\boldsymbol{w})\, d\boldsymbol{w}$.

- In practice, this integral is approximated via Monte Carlo integration by sampling a large number m of directions $\boldsymbol{u}_1, \ldots, \boldsymbol{u}_m$ uniformly on $\mathbb{S}^{d-1}$ and calculating

$$\overline{\mathrm{D}}_{\mathrm{KL}}[f_U \| f_{\mathcal{D}}] := \frac{1}{m}\sum_{i=1}^{m} \log[f_U(\boldsymbol{u}_i)/f_{\mathcal{D}}(\boldsymbol{u}_i)] = -\log[A_{d-1}(1)] - \frac{1}{m}\sum_{i=1}^{m} \log[f_{\mathcal{D}}(\boldsymbol{u}_i)], \tag{7}$$

  with $\overline{\mathrm{D}}_{\mathrm{KL}}[f_U \| f_{\mathcal{D}}] \xrightarrow{\mathbb{P}} \mathrm{D}_{\mathrm{KL}}[f_U \| f_{\mathcal{D}}]$ as $m \to \infty$.

Model assessment

Model assessment – A random point measure approach

- Under assumptions of uniform convergence on $\mathbb{S}^{d-1}$,

$$F_{R|\boldsymbol{W}}\left(\frac{R - r_{\mathcal{Q}_q^\star}(\boldsymbol{W})}{r_{\mathcal{G}^\star}(\boldsymbol{W})}\right)^{1/d}\boldsymbol{W}\Big|\left\{R > r_{\mathcal{Q}_q^\star}(\boldsymbol{W})\right\} \xrightarrow{d} \boldsymbol{U}_{B_1(\boldsymbol{0})}, \quad \text{as } q \to 1,$$

  where $r_{\mathcal{Q}_q^\star}$ and $r_{\mathcal{G}^\star}$ are deterministic functions of $r_{\mathcal{Q}_q}$, $r_{\mathcal{G}}$, and $f_{\boldsymbol{W}}$.

- We consider the stationary random point measure

$$P^\star := \sum_{i=1}^{n} \delta\left[H_{\boldsymbol{W}_i}\left(\frac{R_i - r_{\mathcal{Q}_q^\star}(\boldsymbol{W}_i)}{r_{\mathcal{G}^\star}(\boldsymbol{W}_i)}\right)^{1/d}\boldsymbol{W}_i\right]\mathbb{1}_{R_i > r_{\mathcal{Q}_q^\star}(\boldsymbol{W}_i)}.$$

- We use an adapted version of the standard K-functions to assess if $P^\star$ is statistically distinguishable from a random point measure with constant intensity on $B_1(\boldsymbol{0})$.

Model assessment – A random point measure approach

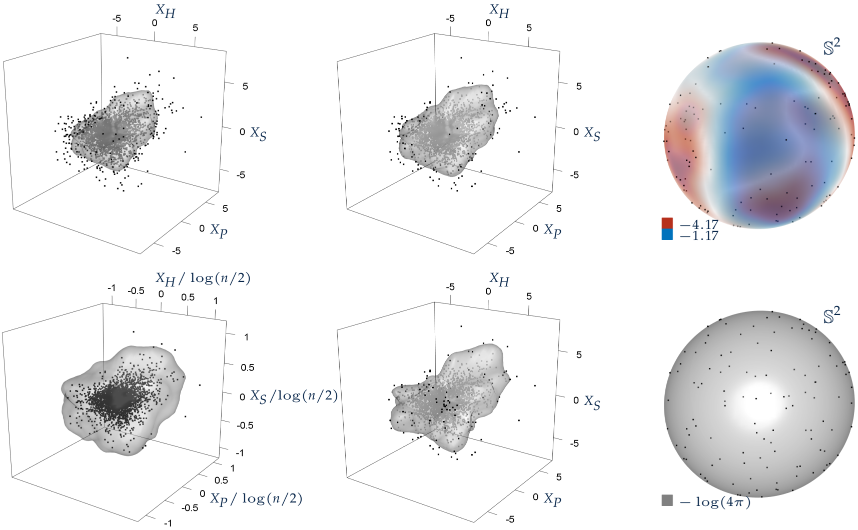


Figure from Papastathopoulos et al. (2023)

Model assessment – A random point measure approach

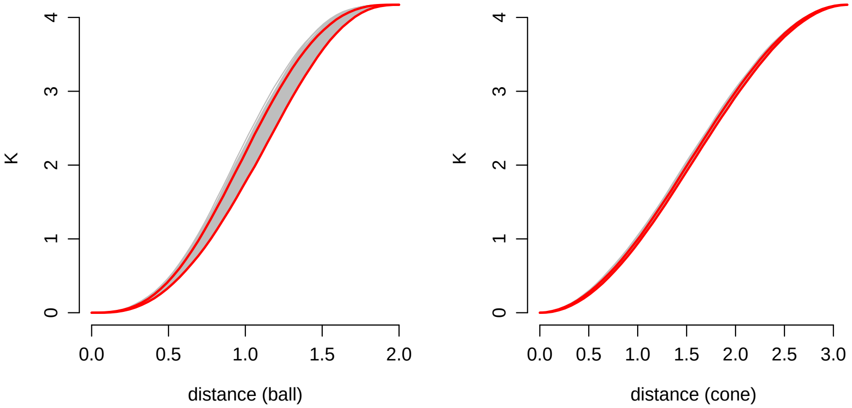
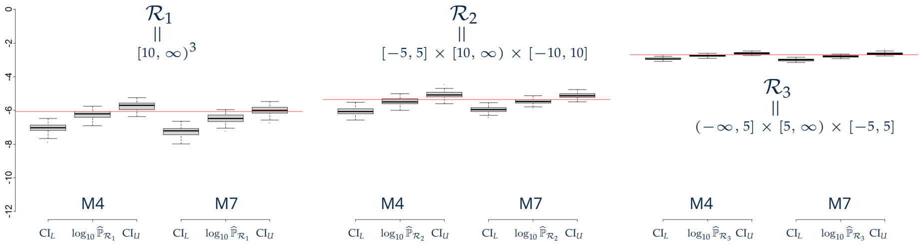
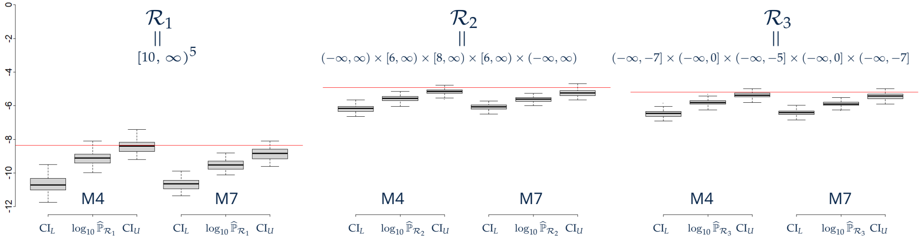


Figure from Papastathopoulos et al. (2023)

Simulation study results – 3 dimensions



Simulation study results – 3 dimensions



Simulation study results – 3 dimensions