Local goodness-of-fit diagnostics for extreme value regression models

Jordan Richards

School of Mathematics and Maxwell Institute for Mathematical Sciences,
University of Edinburgh, UK

22/05/2026

Extremes

Any event, environmental or otherwise, that occurs with **very low** probability.

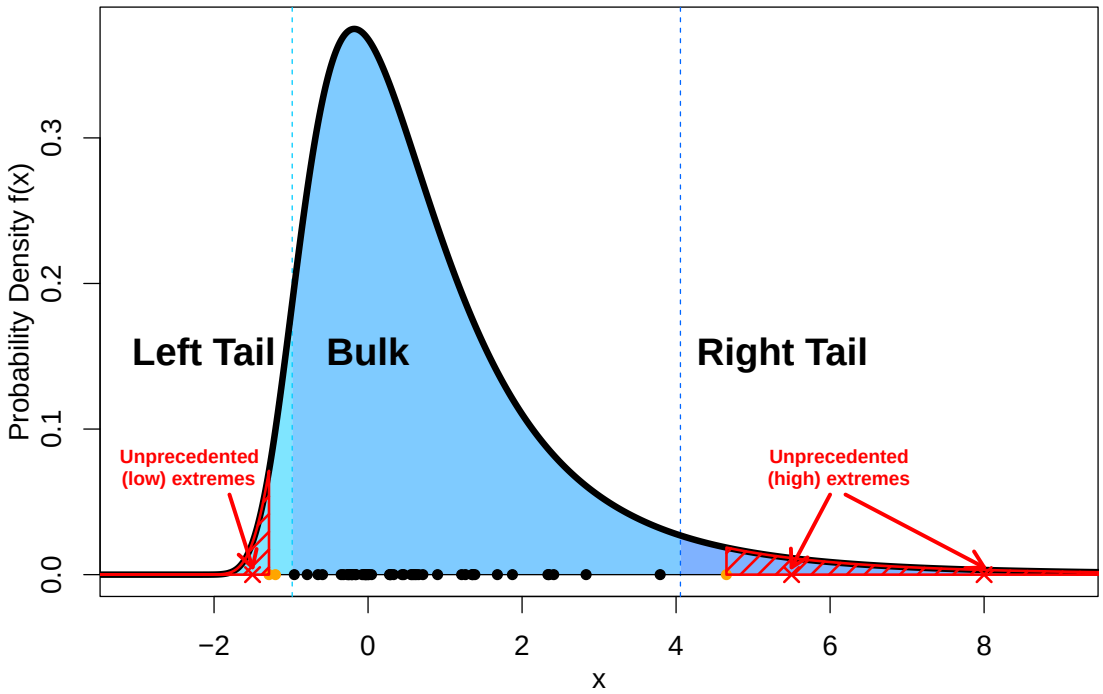



Outline

1. Give a brief introduction to extreme value regression
2. Propose new quantitative and visual **goodness-of-fit diagnostics** for global and regional assessment
   Joint with Ed Mackay and Phil Jonathan
3. **Application:** Deep radial-angular modelling of multivariate extremes
   Joint with Ed Mackay, Callum Murphy-Barltrop, Phil Jonathan

Background on extreme events

- Univariate context: Observations $Y_1, \ldots, Y_n$.
  Can we estimate the probability of unprecedented extreme events of a given size (typically larger than $M_n = \max(Y_1, \ldots, Y_n)$)?



Marginal modeling of extremes — theory for peaks-over-threshold

- **Pickands–de Haan–Balkema Theorem:** high threshold excesses $Y - u \mid Y > u$ may be approximated by the generalized Pareto (GP) distribution, in the sense that there exists a scaling function $a(u) > 0$ such that as $u \to y_F$ (upper endpoint)

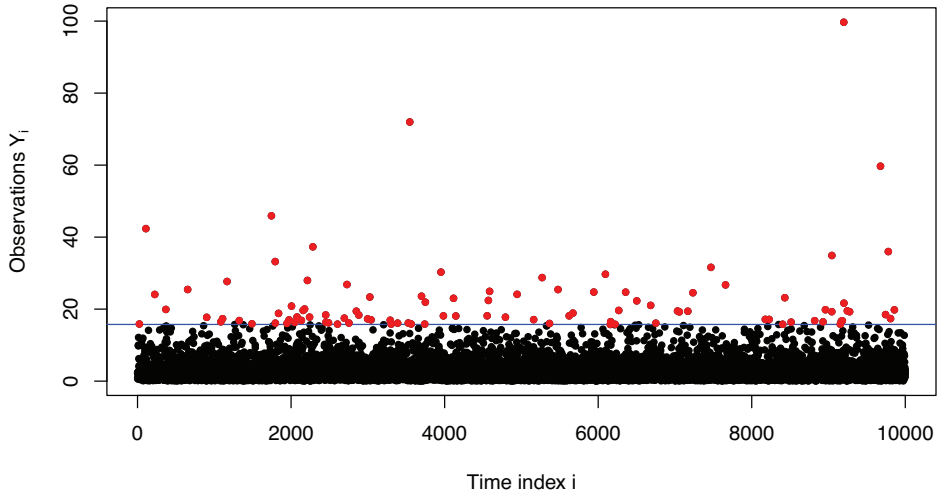
$$\Pr\left(\frac{Y-u}{a(u)} > y \mid Y > u\right) \to 1 - H(y) := \begin{cases} (1+\xi y/\sigma)_+^{-1/\xi}, & \xi \neq 0, \\ \exp(-y/\sigma), & \xi = 0, \end{cases},$$

  where $\sigma > 0$ and $\xi \in \mathbb{R}$ are GP scale and shape parameters, respectively.

- In practice, we can model excesses directly as $Y - u \mid Y > u \sim \mathrm{GP}(\sigma, \xi)$.

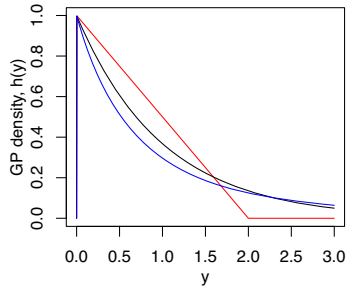
Marginal modeling — Statistics of peaks-over-threshold

Threshold exceedance approach $Y_1, Y_2, \ldots, \overset{\text{iid}}{\sim} F$



Pickands–de Haan–Balkema Theorem: *For a broad range of distributions F, we have the following large-u approximation*

$$Y_i - u \mid Y_i > u \sim \mathrm{GP}(\sigma, \xi).$$



Extreme value regression

Now, introduce data pairs $(Y_i, \mathbf{X}_i)_{i=1}^n$.

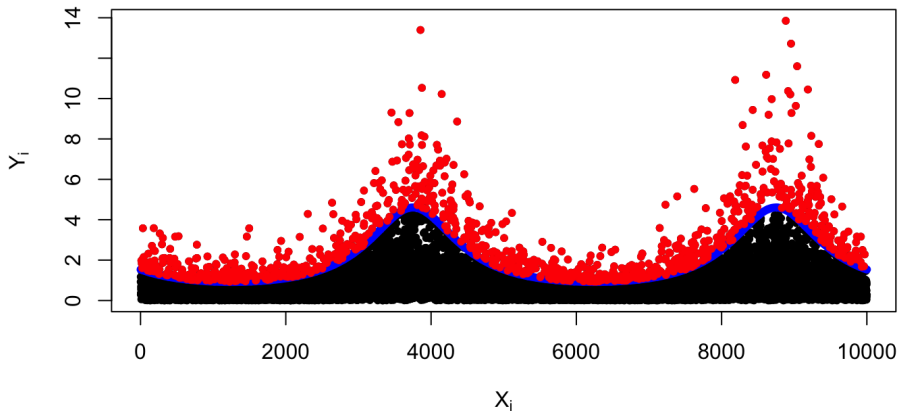
Stats 101

Given Y and covariates $\mathbf{X} \in \mathbb{R}^p$, how do we model the impacts of $\mathbf{X}$ on the extremes of Y?

Answer: **Regression**

In the GP case, we model

$$Y_i - u(\mathbf{x}_i) | (Y_i > u(\mathbf{x}_i), \mathbf{X} = \mathbf{x}_i) \sim \mathrm{GP}(\sigma(\mathbf{x}_\mathrm{i}), \xi(\mathbf{x}_\mathrm{i})).$$



Extreme value regression

To model extremes of $Y \mid \mathbf{X} = \mathbf{x}$, assume $Y \mid \mathbf{X} = \mathbf{x} \sim \mathcal{F}(\boldsymbol{\theta}(\mathbf{x}))$ with parameters $\boldsymbol{\theta} : \mathbb{R}^p \mapsto \mathbb{R}^q$.

Standard choices for $\mathcal{F}$ (in extreme analyses):

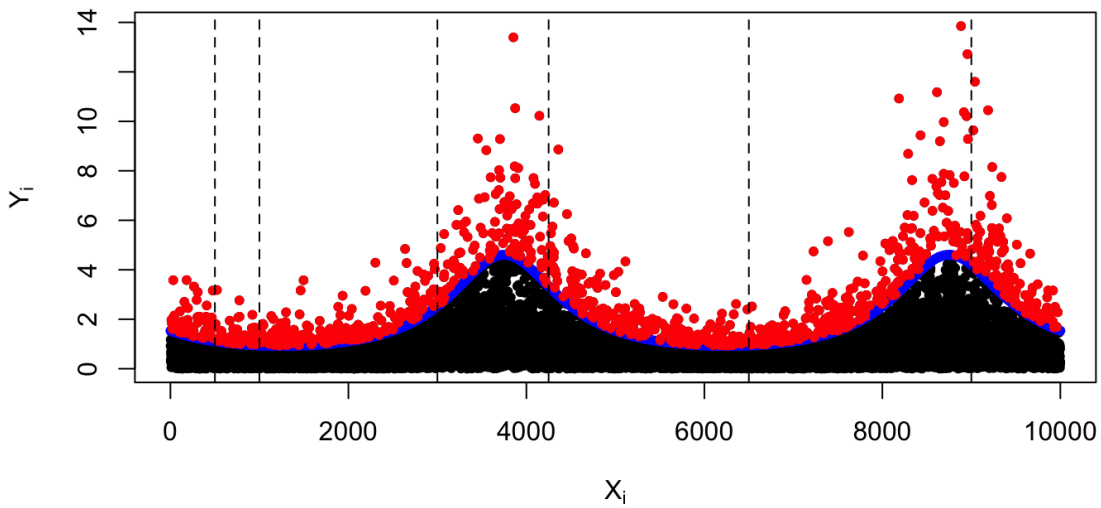
- $\mathcal{F} =$ GP with $\boldsymbol{\theta}(\mathbf{x}) = (\sigma(\mathbf{x}), \xi(\mathbf{x}))$ (for threshold excesses);
- $\mathcal{F} =$ GEV with $\boldsymbol{\theta}(\mathbf{x}) = (\mu(\mathbf{x}), \sigma(\mathbf{x}), \xi(\mathbf{x}))$ (for maxima).
- *Newer* alternatives:
  - eGPD (Papastathopoulos and Tawn, 2013; Cisneros et al., 2024),
  - point processes (Richards and Huser, 2026b),
  - bGEV (Castro-Camilo et al., 2022; Richards and Huser, 2026a),
  - bGP (Majumder and Richards, 2026).

Functional form of $\boldsymbol{\theta}(\mathbf{x})$ is **not important!**

Majumder, Richards (2025). Semi-parametric bulk and tail regression using spline-based neural networks. Extremes.

Motivation

- **Extreme value regression models** arise in many applied domains, e.g., environment, public health, finance, social sciences,...
- Reliable extrapolation requires **careful** assessment of model fit;
- Standard diagnostics pool across the covariate space



Objectives:

1. Assess overall **goodness-of-fit**
2. Identify regions of covariate space with poor fit
3. Facilitate comparison between competing models

Review of visual diagnostics

We have a sample of observations $\{y_1, ..., y_n\}$ of random variable $Y \sim F_Y$:

- Used to estimate a model with distribution function $\hat{F}_Y$;
- (PIT) if $\hat{F}_Y = F_Y$, the values $\{\hat{F}_Y(y_1), ..., \hat{F}_Y(y_n)\} \sim \text{Uniform}(0, 1)$.

Let $y_{(1)} \leq \cdots \leq y_{(n)}$ denote the ordered sample:

- assign each $y_{(k)}$ an empirical non-exceedance probability p_k
- definition of p_k varies between practitioners, with $p_k = (k - 0.5)/n$ or $p_k = k/(n + 1)$ being common choices.

Review of visual diagnostics

A quantile-quantile (QQ) plot consists of the points

$$\left\{ \left(\hat{F}_Y^{-1}(p_k),\, y_{(k)} \right) : k = 1, ..., n \right\},$$

where $\hat{F}_Y^{-1}$ is the model's quantile function.

A transformed QQ plot consists of the points

$$\left\{ \left(F_0^{-1}(p_k),\, F_0^{-1}\left(\hat{F}_Y(y_{(k)}) \right) \right) : k = 1, ..., n \right\}.$$

In classical statistics, F_0 often Φ.

Non-stationary visual diagnostics

In a regression setting, we instead have samples $\{(y_i, \mathbf{x}_i)\}_{i=1}^n$, where Y is now conditional on $\mathbf{X} \in \mathcal{D} \subset \mathbb{R}^p$

- Samples used to estimate a conditional model $\hat{F}_{Y|X}$;
- (PIT) if $\hat{F}_{Y|\mathbf{X}} = F_{Y|\mathbf{X}}$, the independent values $\{\hat{F}_{Y|\mathbf{X}}(y_1|\mathbf{x}_1), ..., \hat{F}_{Y|\mathbf{X}}(y_n|\mathbf{x}_n)\} \sim \text{Uniform}(0, 1)$.

We can adapt transformed QQ plots to the non-stationary setting:

$$\left\{ \left(F_0^{-1}(p_k),\, F_0^{-1}\left(\hat{F}_{Y|\mathbf{X}}(y_{(k)}|\mathbf{x}_{(k)})\right)\right) : k = 1, ..., n \right\},$$

where $\mathbf{x}_{(k)} = \{\mathbf{x}_{(i)} : y_i = y_{(k)}\}$.

Pooled exponential QQ plot

Define $P_k = F_{Y|\boldsymbol{X}}(Y_k|\boldsymbol{X}_k)$, $Q_k = 1 - P_k$, and $Z_k = -\log(Q_k)$. Since $P_k \sim \text{Uniform}(0,1)$, we have $Z_k \sim \text{Exp}(1)$.

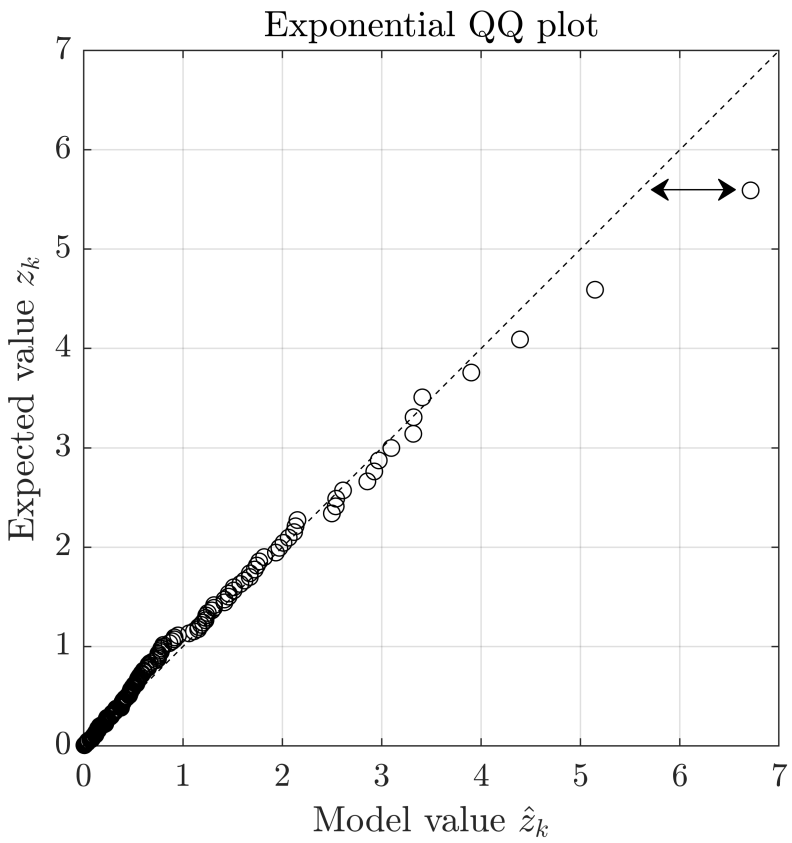
Denote the **ordered exceedance probabilities** $Q_{(1)} \leq \cdots \leq Q_{(n)}$ and exponential order statistics $Z_{(1)} \geq \cdots \geq Z_{(n)}$.

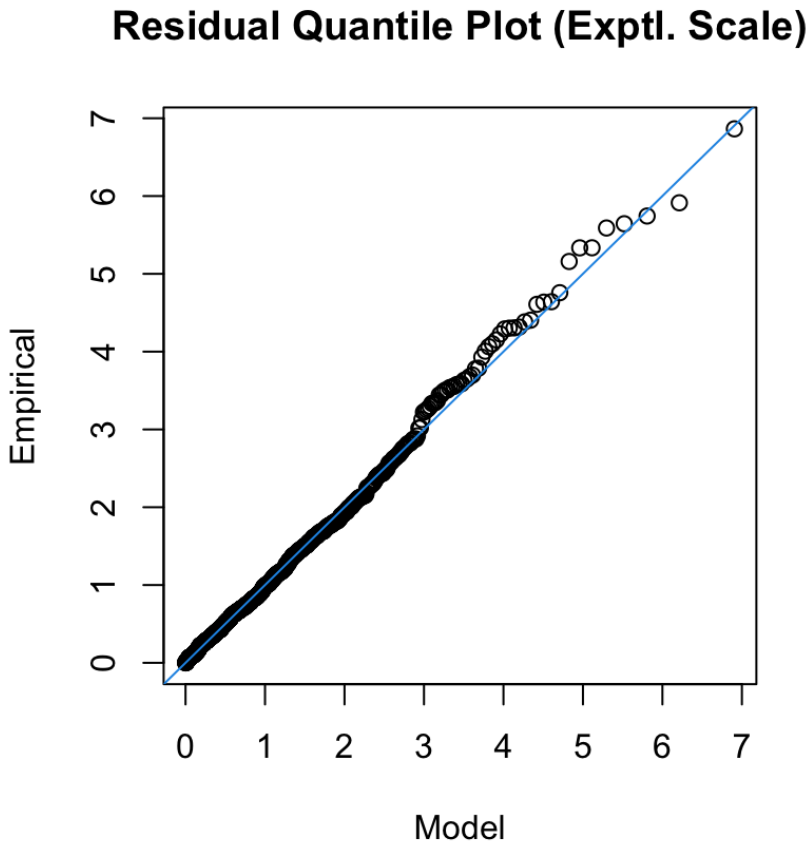
For a sample of observations $\{(y_i, \boldsymbol{x}_i)\}_{i=1}^n$ and a model $\hat{F}_{Y|\boldsymbol{X}}$, define $\hat{q}_k = 1 - \hat{F}_{Y|\boldsymbol{X}}(y_k|\boldsymbol{x}_k)$. Denote the ordered values $\hat{q}_{(1)} \leq \cdots \leq \hat{q}_{(n)}$, and define $\hat{z}_k = -\log(\hat{q}_{(k)})$. The pooled exponential QQ plot is a scatter plot of

$$\{(\hat{z}_k, z_k) : k = 1, ..., n\},$$

where z_k is (in our case) $\mathbb{E}[Z_{(k)}]$.

Pooled exponential QQ plot

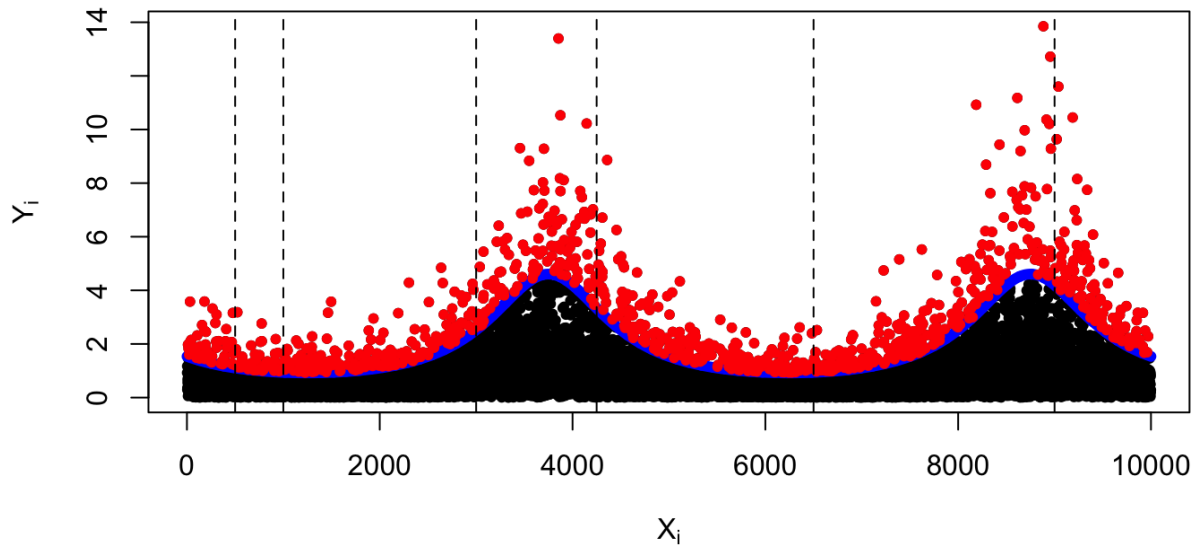




Early use: Heffernan, J.E., Tawn, J.A. Extreme Value Analysis of a Large Designed Experiment: A Case Study in Bulk Carrier Safety. Extremes 4, 359–378 (2001).

Limitation of Exponential QQ Plots

Main issue - no local/regional information:



Suppose now $\mathcal{D}$ is partitioned into B non-overlapping regions, or 'bins', $\mathcal{B}_1, ..., \mathcal{B}_B$ such that $\cup_{b=1}^{B} \mathcal{B}_b = \mathcal{D}$ and $\cap_{b=1}^{B} \mathcal{B}_b = \emptyset$, with $n_b = |\mathcal{B}_b|$.

Local exponential QQ plots

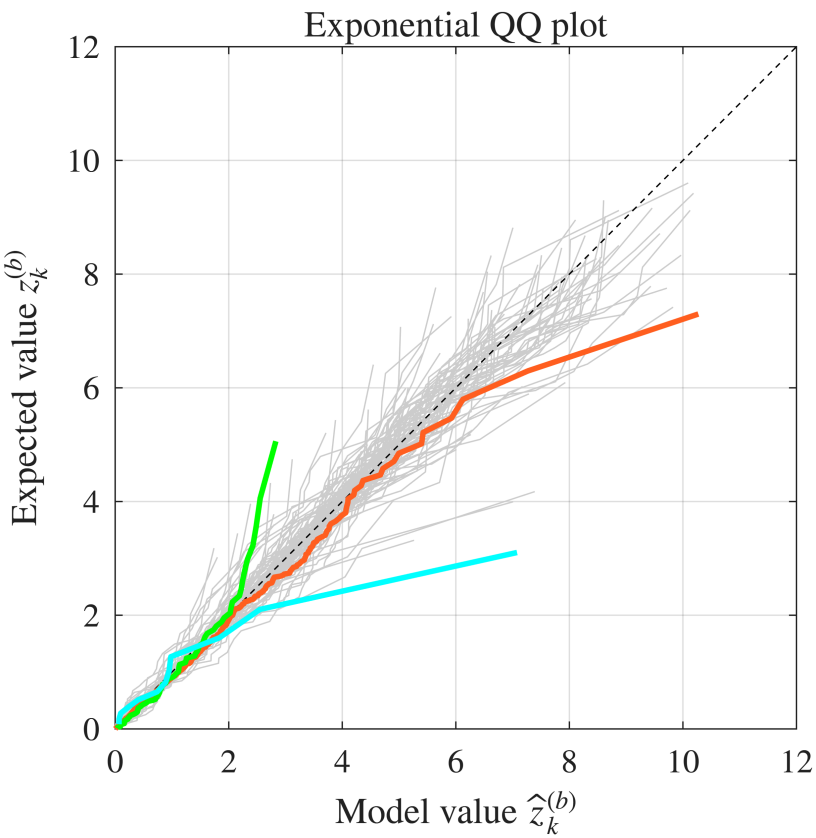
First idea: define bin-wise $Z_k^{(b)}$ (and similar). For bin $b \in \{1, \ldots, B\}$, define $\mathbf{X}_k^{(b)} \in \mathcal{B}_b$ for $k = 1, \ldots, n_b$. Then,

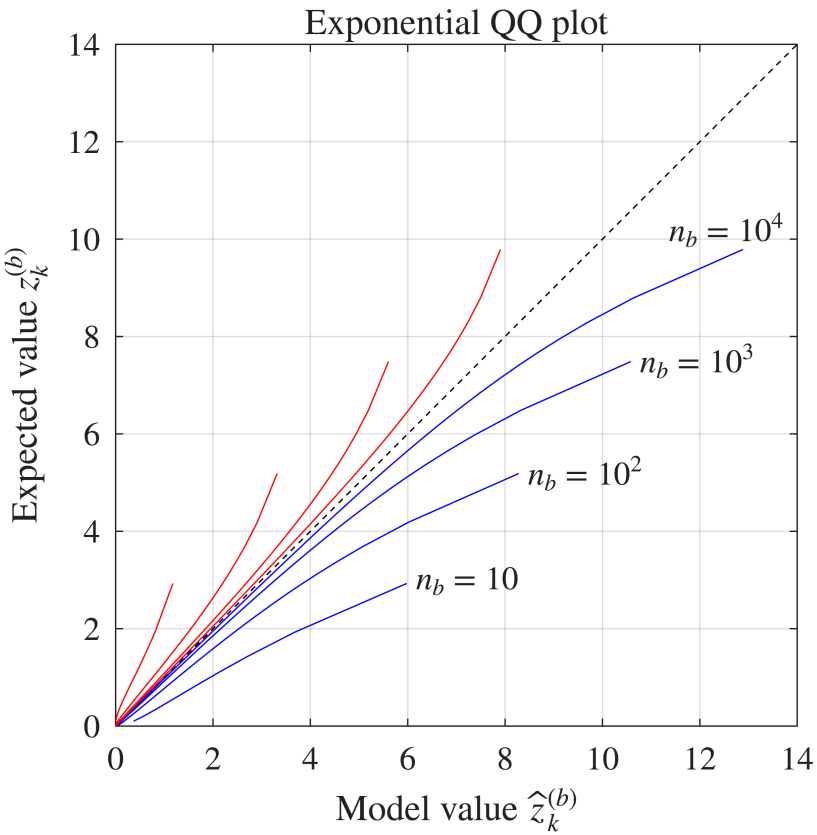
$$Z_k^{(b)} = -\log(Q_k^{(b)}) = -\log\{1 - F_{Y|\mathbf{X}}(Y_k^{(b)}|\mathbf{X}_k^{(b)})\}.$$

Evaluate **pooled exponential QQ plot** for each bin, $\mathcal{B}_1, \ldots, \mathcal{B}_B$, with $\hat{F}_{Y|\mathbf{X}}$ constant across bins. Overlay on the same graph.

Unfortunately, they are not invariant to sample size n_b, making consistent goodness-of-fit and uncertainty assessment infeasible.

Limitation of exponential QQ plots





Standardised exponential residuals

For $k = 1, \ldots, n_b$, define exponential residuals

$$D_k = \mathbb{E}\left[Z_{(k)}\right] - Z_{(k)},$$

as the **difference** between the expected and observed k-th exponential order statistic.

Closed form:

$$z_k := \mathbb{E}\left[Z_{(k)}\right] = H_n - H_{k-1}, \tag{1}$$

where $H_k = \sum_{j=1}^{k} \frac{1}{j}$ is the k-th harmonic number, and we define $H_0 = 0$.

Corollary

For any fixed $k \in \mathbb{N}_{>0}$, D_k converges in distribution, as $n \to \infty$, to a log-gamma random variable, $D_{k,\infty}$, with shape parameter k, unit scale, and location $\mu_k = \gamma - H_{k-1}$.

Standardised exponential residuals

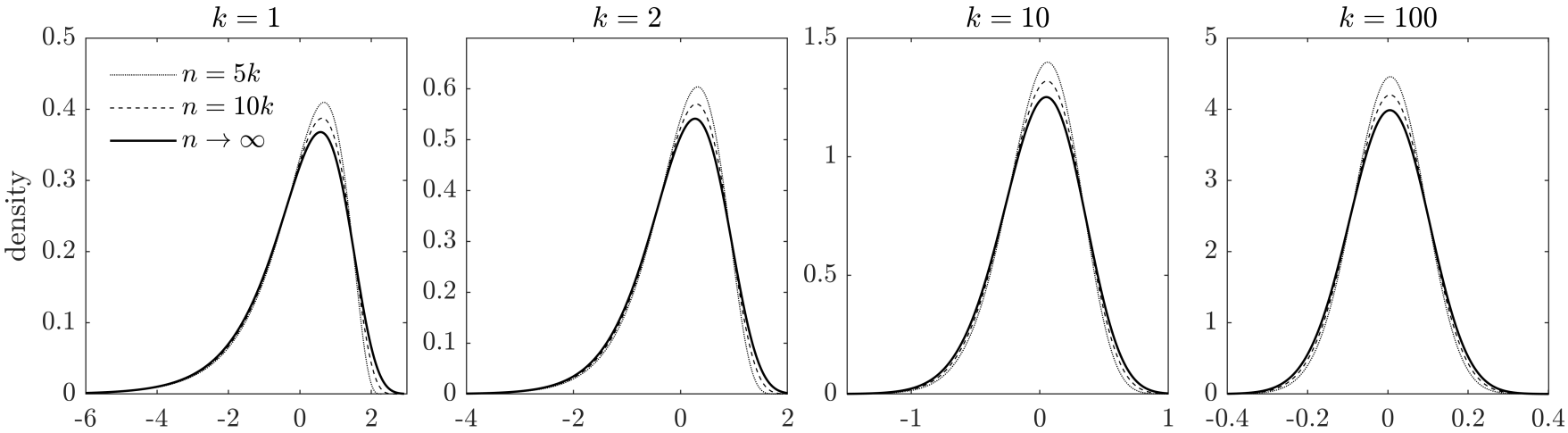


Figure: Densities of $D_k = \mathbb{E}\left[Z_{(k)}\right] - Z_{(k)}$, for various ranks k and sample sizes n.

The density function of $D_{k,\infty}$ is

$$f_{D_{k,\infty}}(x) = \frac{1}{\Gamma(k)} \exp\left[k\left(x - \mu_k\right) - \exp\left(x - \mu_k\right)\right], \quad x \in \mathbb{R}, \qquad (2)$$

where Γ is the gamma function.

Fact Convergence - **Implication**: Sampling distribution of extremes is approximately independent of n_b.

Standardised Tail Plot

We define the regional *standardised tail plot* to be lines joining the pairs

$$\{(k, z_k^{(b)} - \hat{z}_k^{(b)}) : k = 1, \ldots, n_b\}, \tag{3}$$

where $z_k^{(b)} = \mathbb{E}[Z_{(k)}^{(b)}]$ for each bin $\mathcal{B}_1, ..., \mathcal{B}_B$ in the covariate domain.

Standardised Tail Plot

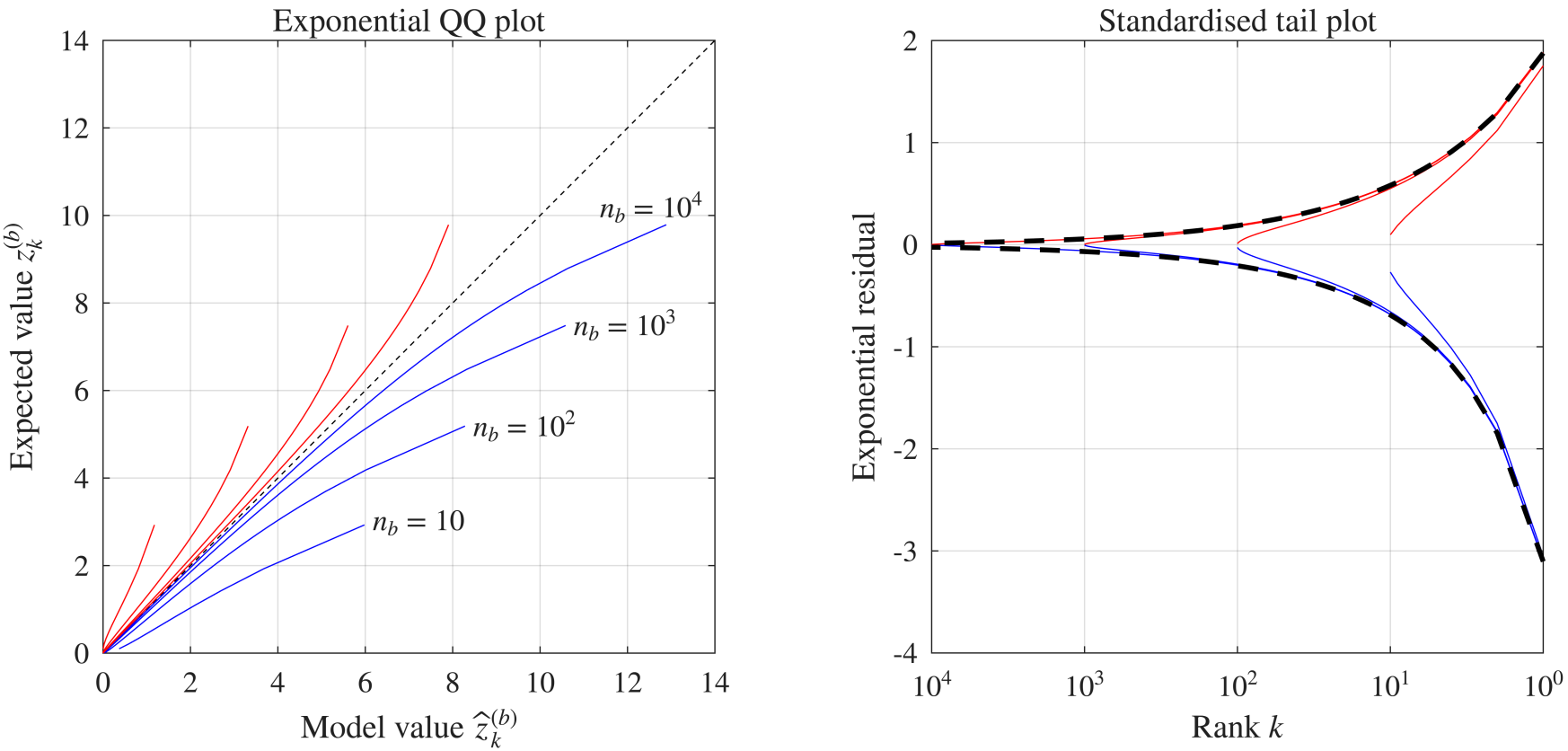


Figure: Left: Two-sided 95% CI on exponential quantiles for sample sizes of $n_b = 10, 10^2, 10^3, 10^4$. Right: Two-sided 95% CI for differences D_k between model and expected exponential quantiles for the same sample sizes, together with 95% CI for the asymptotic distribution (black dashed lines).

Standardised Tail Plot

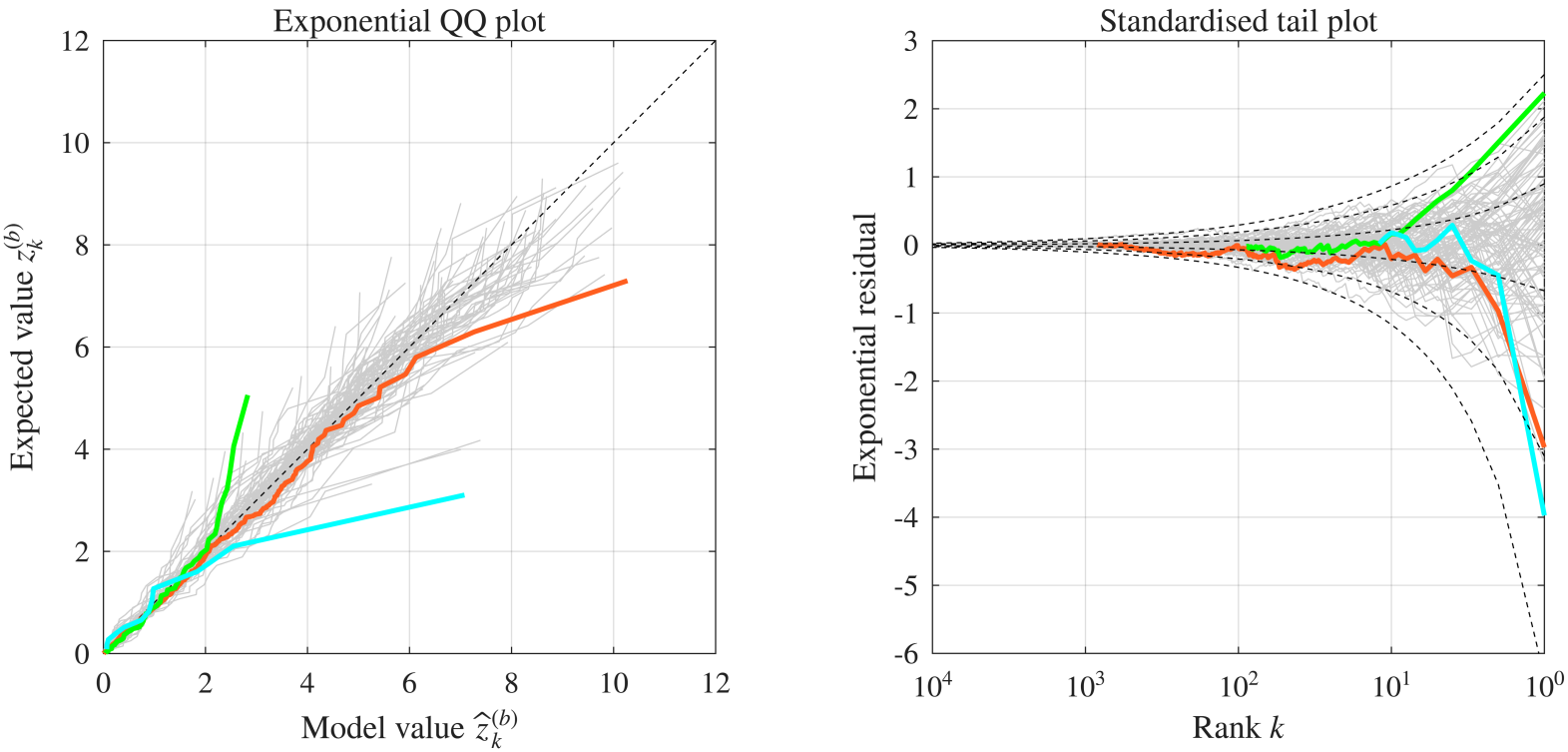


Figure: Left: Example simulations for 100 samples of random size N, where $\log_{10}(N) \sim \mathrm{U}(1, 4)$. Right: Same data transformed to standardised scale, together with quantiles of the asymptotic distribution at non-exceedance probabilities 0.001, 0.025, 0.25, 0.75, 0.975, and 0.999 (dashed lines).

Standardised Tail Plot

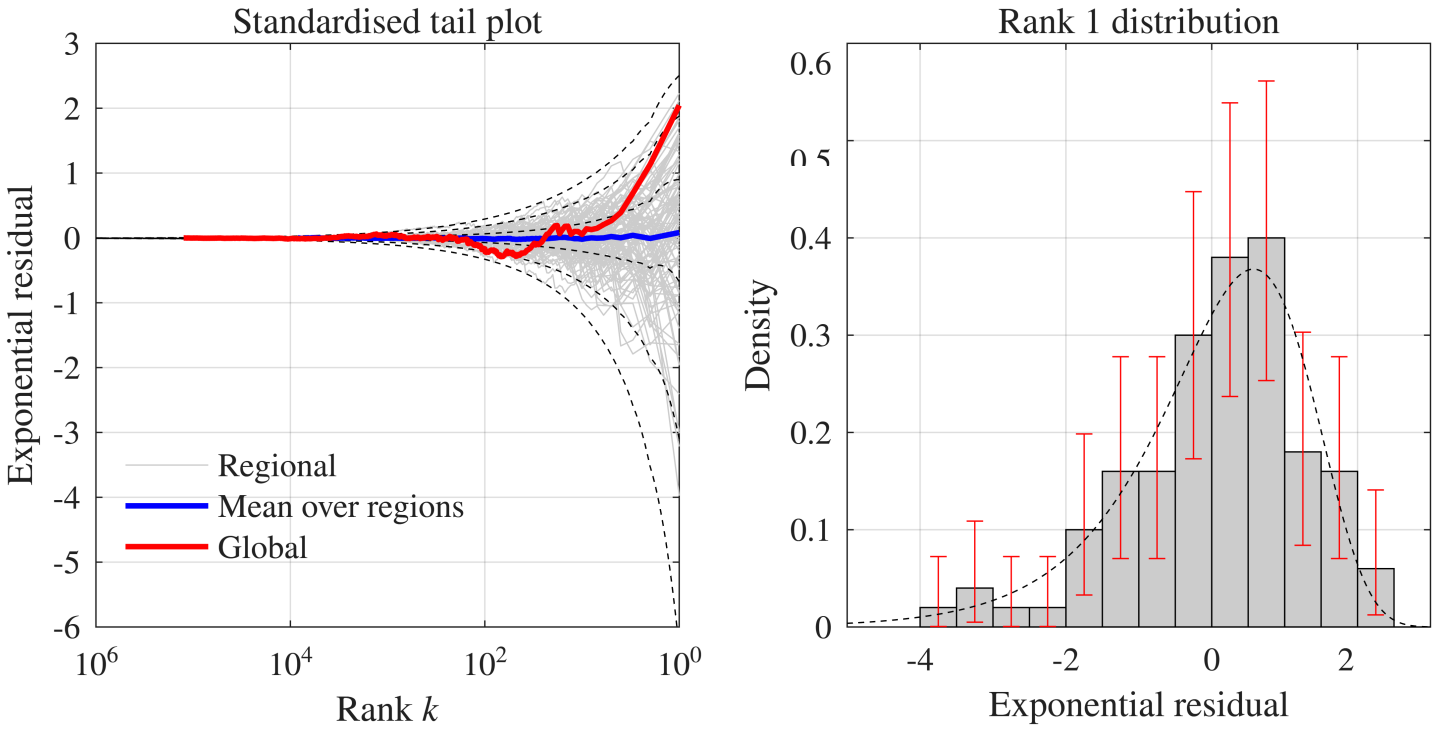


Figure: Bold blue lines are the mean value at a given rank k, over all bins. Red lines in these plots are the corresponding global diagnostic. Right: density histogram of the rank $k = 1$ exponential residuals over all regions, together with 95% error bars. The asymptotic sampling distribution is shown as the dashed line.

Diagnostic plots - comments

Standardised tail plot:

- Facilitates visualisation of goodness-of-fit consistently; both i) globally and across multiple regions of the covariate space;
- Standardisation allows for (approximately) sample-size-independent confidence bounds;
- Reads in the same way as traditional pooled exponential QQ plots;
- We also create a **normal** version if interested in bulk fit as well!

Summary statistics

For quantitative summaries, traditional analyses use the Cramér–von Mises (CvM) family of goodness-of-fit statistics. We can define these both globally and regionally.

Let $\mathbb{F}_n(u)$ by the EDF of model PIT values, $\hat{q}_1 \leq \cdots \leq q_n$. The CvM family of test statistics is defined as

$$T_n^2 = n \int_0^1 \psi(u) \left(\mathbb{F}_n(u) - u\right)^2 \mathrm{d}u,$$

where $\psi : [0,1] \to [0,\infty)$ is some non-negative weight function.

Under the null hypothesis - $H_0 : \hat{F}_{Y|\mathbf{X}} = F_{Y|\mathbf{X}}$:

$$T_n^2 \xrightarrow{d} T^2 = \int_0^1 \psi(u) \left(B(u)\right)^2 \mathrm{d}u, \tag{4}$$

where B is a Brownian bridge on $[0,1]$.

Summary statistics

Popular choices:

$$\begin{aligned} \text{Cramer von-Mises (CVM):} \quad & W_n^2 : \psi(u) = 1, \\ \text{Anderson Darling (AD):} \quad & A_n^2 : \psi(u) = (u(1-u))^{-1}, \\ \text{Anderson Darling Right-tail weighted (ADR):} \quad & A_{R,n}^2 : \psi(u) = u^{-1}. \end{aligned}$$

For ADR, under the null:

$$A_{R,n}^2 \xrightarrow{d} A^2 = \int_0^1 \frac{(B(u))^2}{u} \mathrm{d}u. \tag{5}$$

MAD of exponential order statistics (EMAD)

We propose the EMAD, which is the mean absolute deviation of exponential order statistics:

$$S_n = \frac{1}{\sqrt{n}} \sum_{k=1}^{n} |z_k - \hat{z}_k| \, .$$

Proposition (Asymptotic distribution of EMAD)

Under the null hypothesis - $H_0 : \hat{F}_{Y|\mathbf{x}} = F_{Y|\mathbf{x}}$:

$$S_n \xrightarrow{d} S = \int_0^1 \frac{|B(u)|}{u} \mathrm{d}u, \quad n \to \infty,$$

where B is a Brownian bridge on $[0, 1]$. The asymptotic expectation and variance are $\mathbb{E}[S] = \sqrt{\pi/2}$ and $Var(S) = 4\log(2) - \frac{\pi}{2} - 1$, respectively.

Asymptotic sensitivity

Under H_0, the PIT values are uniformly distributed on $[0, 1]$, with distribution function $F_0(u) = u$, $u \in [0, 1]$.

Suppose we test the distribution F_0 under the null against a sequence of local alternatives that differ by $O(n^{-1/2})$, defined as $F_\theta(u) = u + \theta h(u)$, where $\theta = \lambda/\sqrt{n}$, $\lambda > 0$, and $h(u)$ is a local deviation function.

Asymptotic sensitivity

Denote the CvM family test statistic under H_0 as $T^2_{n,0}$ and the statistic under the local alternative F_θ as $T^2_{n,\theta}$. Then, the asymptotic bias converges to

$$\mathbb{E}[T^2_{n,\theta}] - \mathbb{E}[T^2_{n,0}] \to \lambda^2 \int_0^1 \psi(u)\{h(u)\}^2 \mathrm{d}u. \tag{6}$$

The quantity

$$\delta(u) := \frac{\psi(u)}{\sqrt{\mathrm{Var}(T^2_0)}},$$

defines a local signal-to-noise ratio for perturbations to the PIT distribution at exceedance probability u. For the CVM, this is determined by $\psi(u)$.

Asymptotic sensitivity

Proposition (Asymptotic sensitivity of EMAD)

For small perturbations with $\lambda \to 0$, the asymptotic bias of the EMAD statistic for local alternatives converges, as $n \to \infty$, to

$$\mathbb{E}[S_{n,\theta}] - \mathbb{E}[S_{n,0}] \to \lambda^2 \int_0^1 \frac{h^2(u)}{u\sqrt{2\pi u(1-u)}} \mathrm{d}u.$$

We can thus define a directly comparable asymptotic signal-to-noise ratio for the EMAD statistic as $\delta(u) = \left(u\sqrt{2\pi \mathrm{Var}(S_0) u(1-u)}\right)^{-1}$. So, the EMAD statistic has sensitivity $O(u^{-3/2})$ as exceedance probability $u \to 0$, whereas the ADR statistic has sensitivity $O(u^{-1})$.

Asymptotic sensitivity

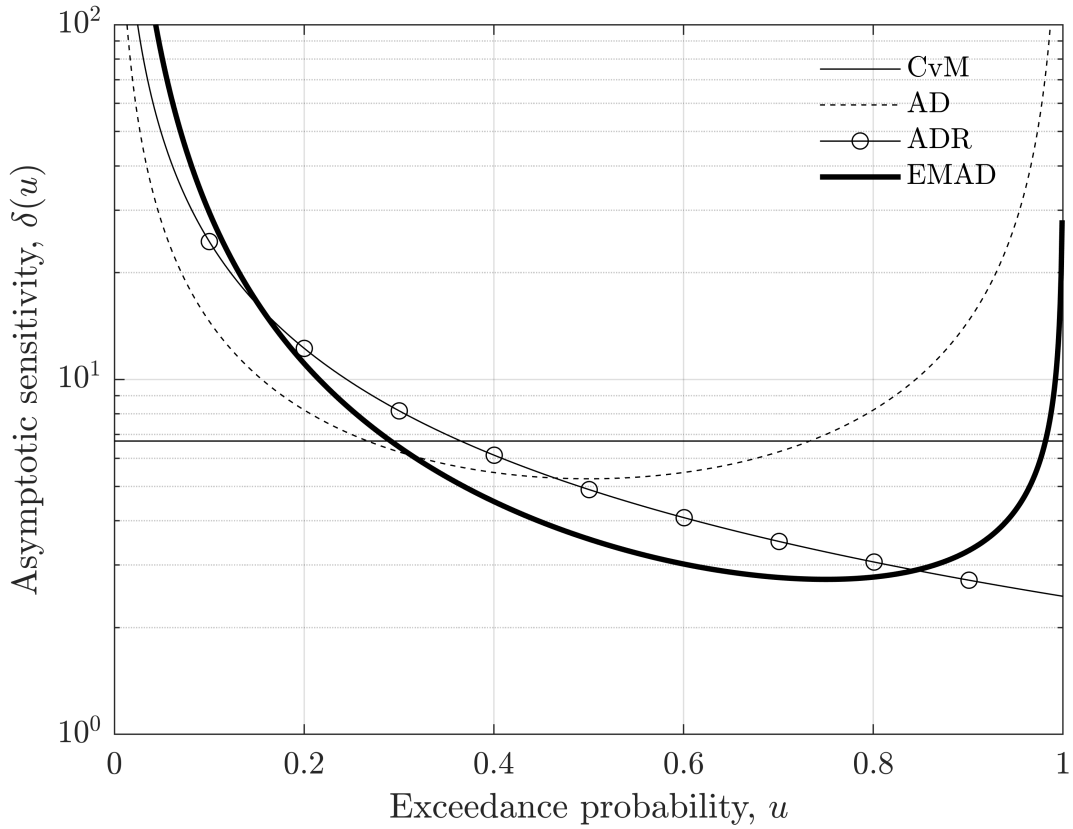
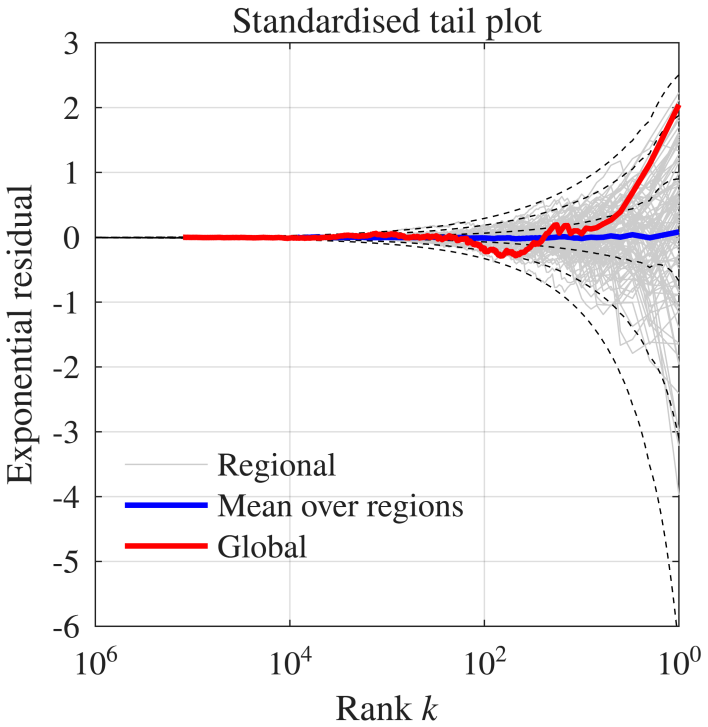


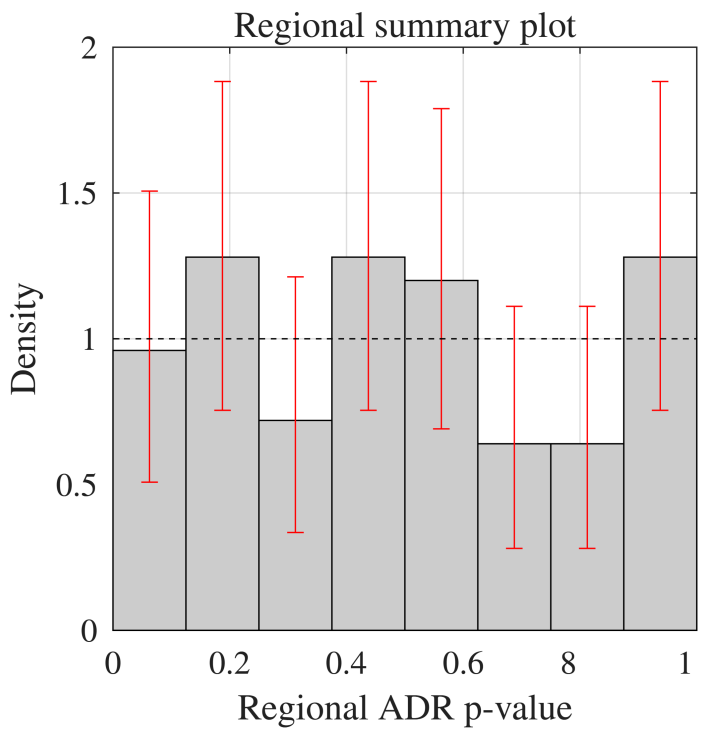
Figure: Asymptotic sensitivity of various goodness-of-fit statistics to perturbations in the PIT distribution at various exceedance probabilities.

Summary statistics

Summary statistics can be evaluated globally or at a regional level:

- If evaluated at a regional level, e.g., $S_{n_b}, b = 1, \ldots, B$, we can summarise across bins by standardising to the associated p-values;
- p-values can be estimated via Monte Carlo methods;
- Across bins, p-values should be iid Uniform(0,1) - this can be further summarised into a single p-value, if doing model comparison!





Summary statistics

To summarise:

1. New visual diagnostic - the **standardised tail plot**;
2. Goodness-of-fit can be quantified regionally or globally used test statistics;
3. Uniformity of regional goodness-of-fit can be visualised, or quantified (using, e.g., CvM test);
4. Goodness-of-fit metrics can be pooled over a bin b (for fixed rank k), over a rank k (for fixed bin b), or across bins and ranks (globally).

Application: Deep SPAR Model

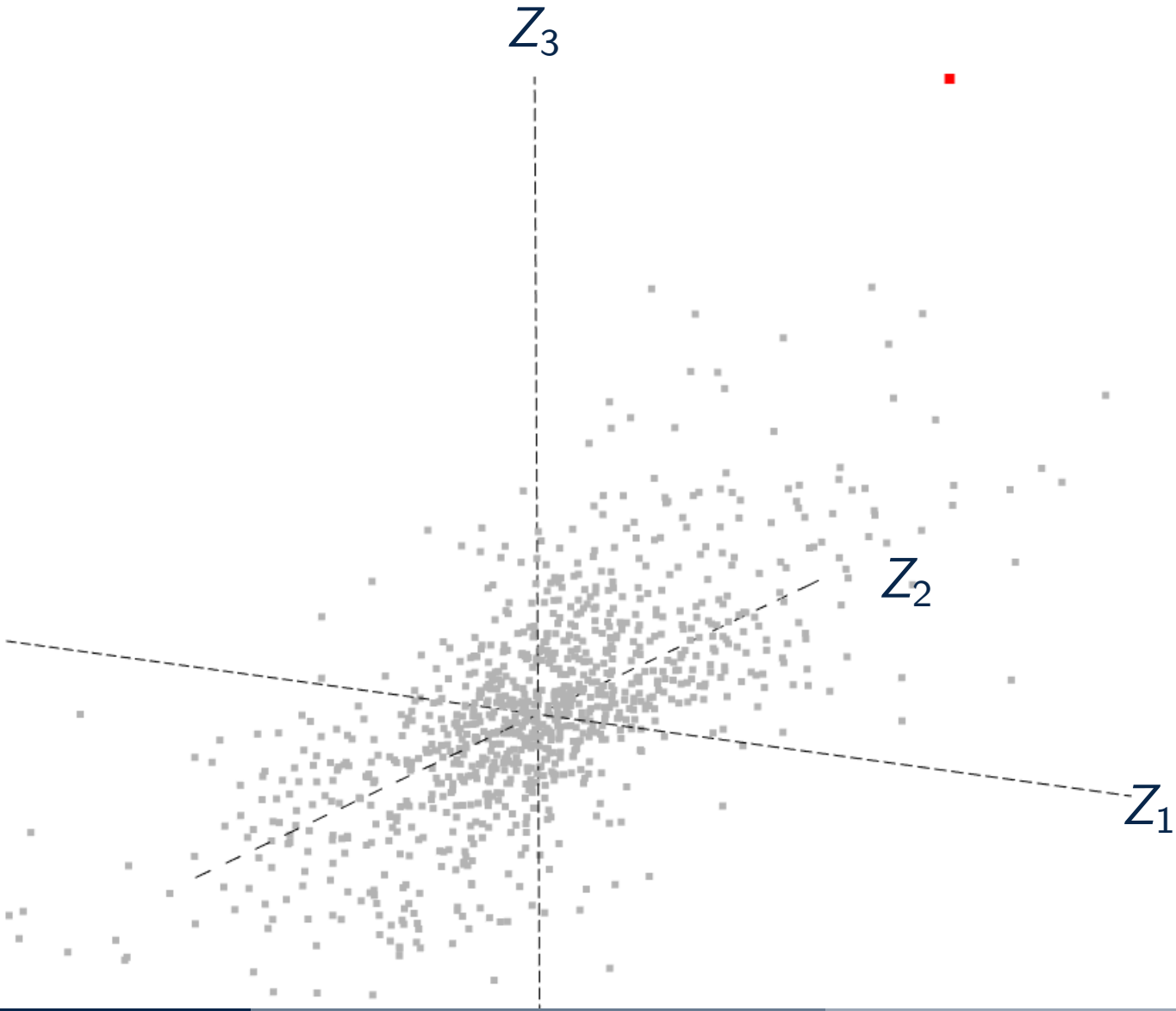
For **multivariate extremes** of RV: $\mathbf{Z} \in \mathbb{R}^d$.
Let

$$Y := \|\mathbf{Z}\| > 0, \qquad \mathbf{X} := \frac{\mathbf{Z}}{\|\mathbf{Z}\|} \in \mathcal{S}^{d-1}.$$

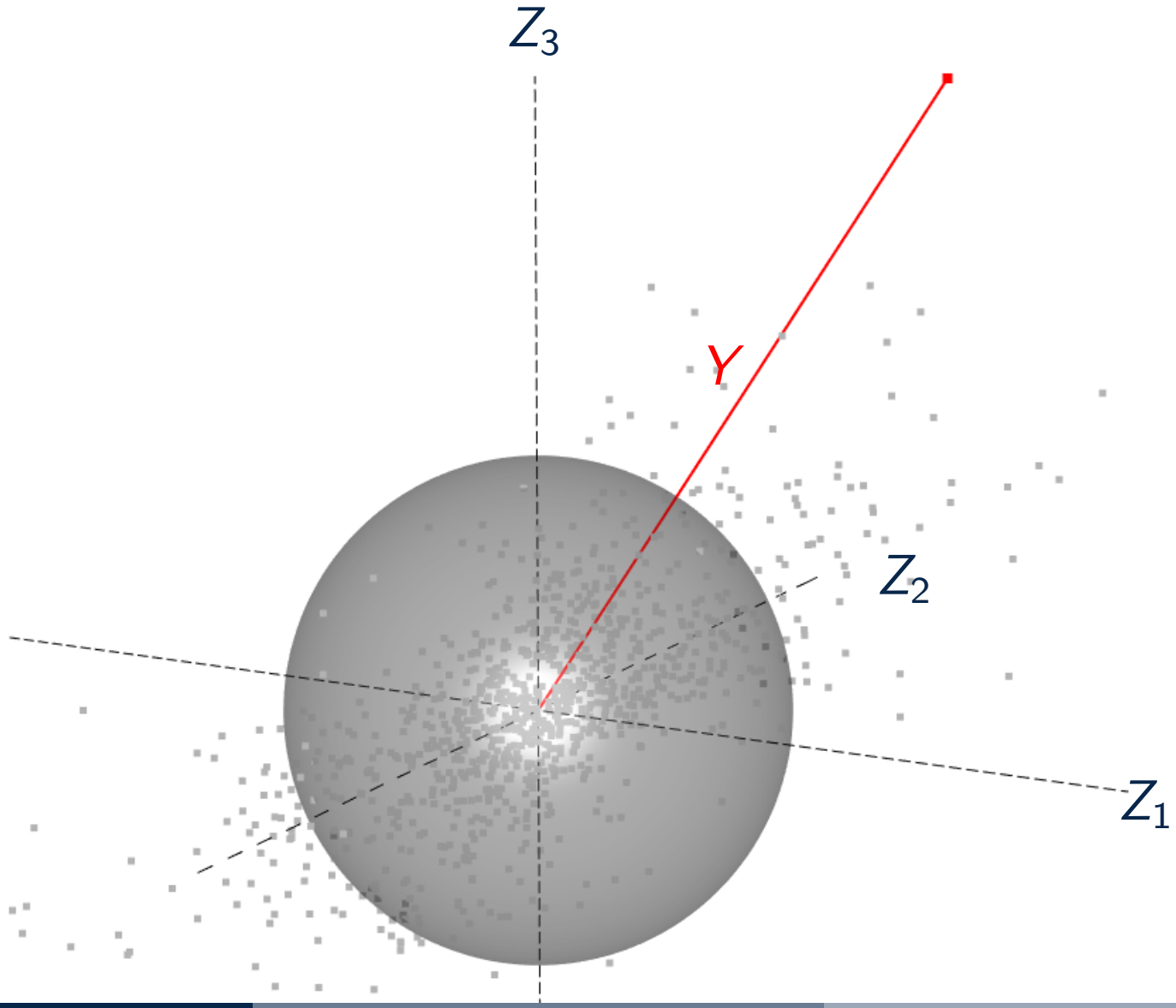
In our application: 20000 training samples for model estimation, from a $d = 5$ Gaussian copula for $\mathbf{Z}$ (Laplace margins).

All diagnostics are evaluated on an independent test set $n = 50000$.

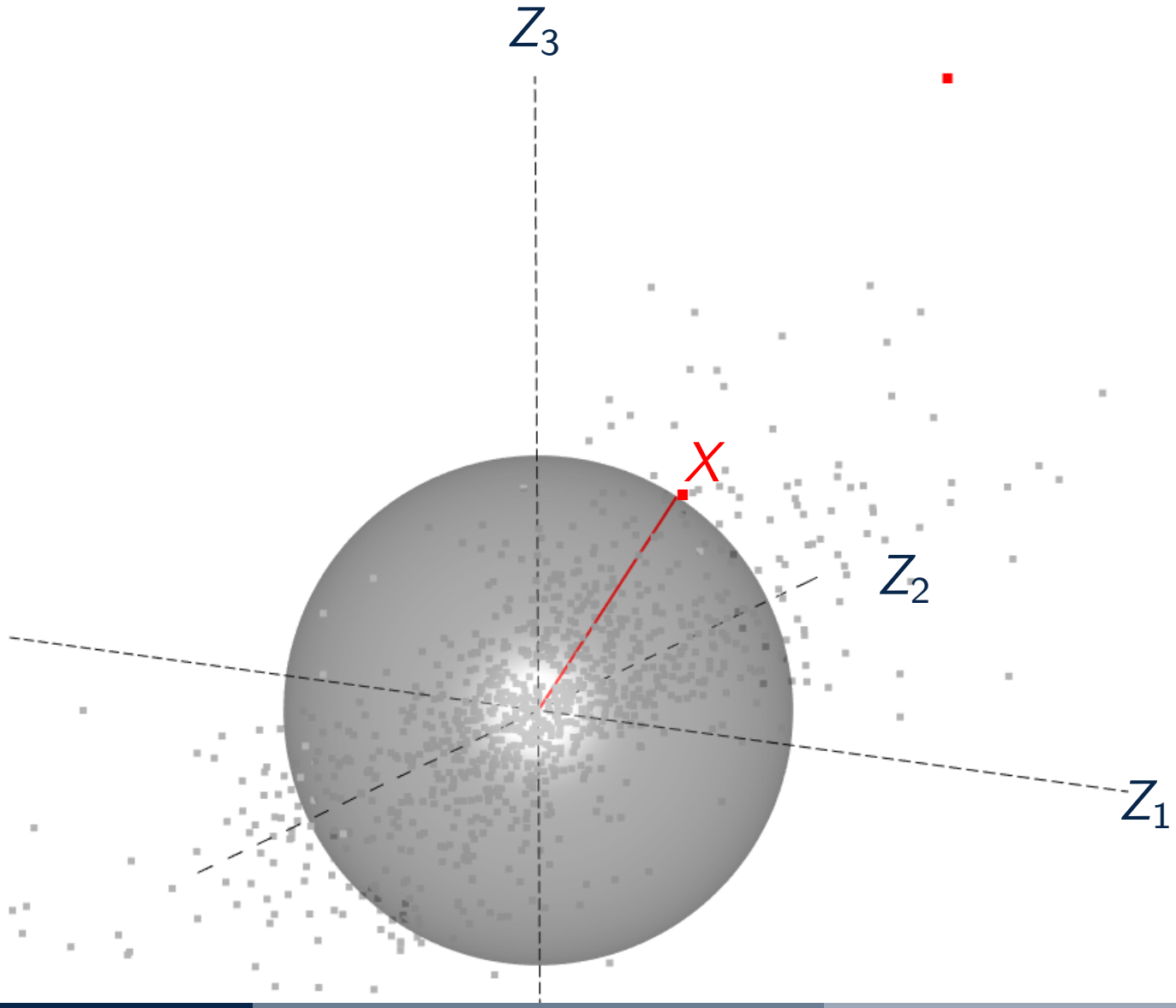
Application: Deep SPAR Model



Application: Deep SPAR Model



Application: Deep SPAR Model



Application: Deep SPAR Model

The joint density of $(Y, \mathbf{X})$ can be written as

$$f_{Y,\mathbf{X}}(y, \mathbf{x}) = f_{\mathbf{X}}(\mathbf{x}) f_{Y|\mathbf{X}}(y|\mathbf{x}). \tag{7}$$

The Semi-Parametric Angular-Radial model (SPAR; Mackay and Jonathan, 2023) replaces $f_{Y|\mathbf{X}}(y|\mathbf{x})$ with a conditional GP model **when** y **is big**, i.e.,

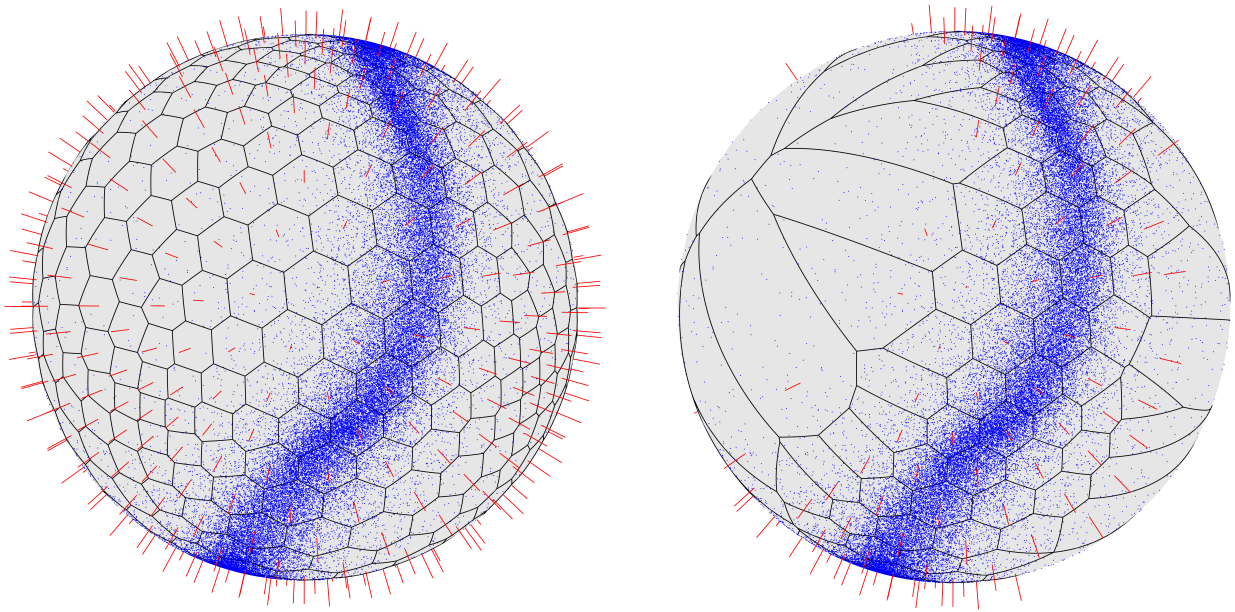
$$f_{Y|\mathbf{X}}(y|\mathbf{x}) \approx (1-\tau) f_{\mathrm{GP}}(y - u(\mathbf{x}); \sigma(\mathbf{x}), \xi(\mathbf{x})), \tag{8}$$

for large $y > u(\mathbf{x})$ and where $\tau := \Pr(Y \leq u(\mathbf{x})|\mathbf{X} = \mathbf{x})$ is fixed.

Mackay, Jonathan (2023). Modelling multivariate extremes through angular-radial decomposition of the density function.

Example - Deep SPAR Model

- Mackay et al. (2025) models $(u(\mathbf{x}), \sigma(\mathbf{x}), \xi(\mathbf{x}))$ via **deep extreme value regression**:
  - 1000s of candidate models to pick from;
  - Threshold level τ needs selecting first;
  - Covariate domain $\mathcal{D}$ is the 4-sphere $\Rightarrow$ $B = 360$ bins chosen via Voronoi partition.



Mackay, Murphy-Barltrop, Richards, Jonathan (2025). Deep Learning Joint Extremes of Metocean Variables Using the SPAR Model. JOMAE.

Threshold selection

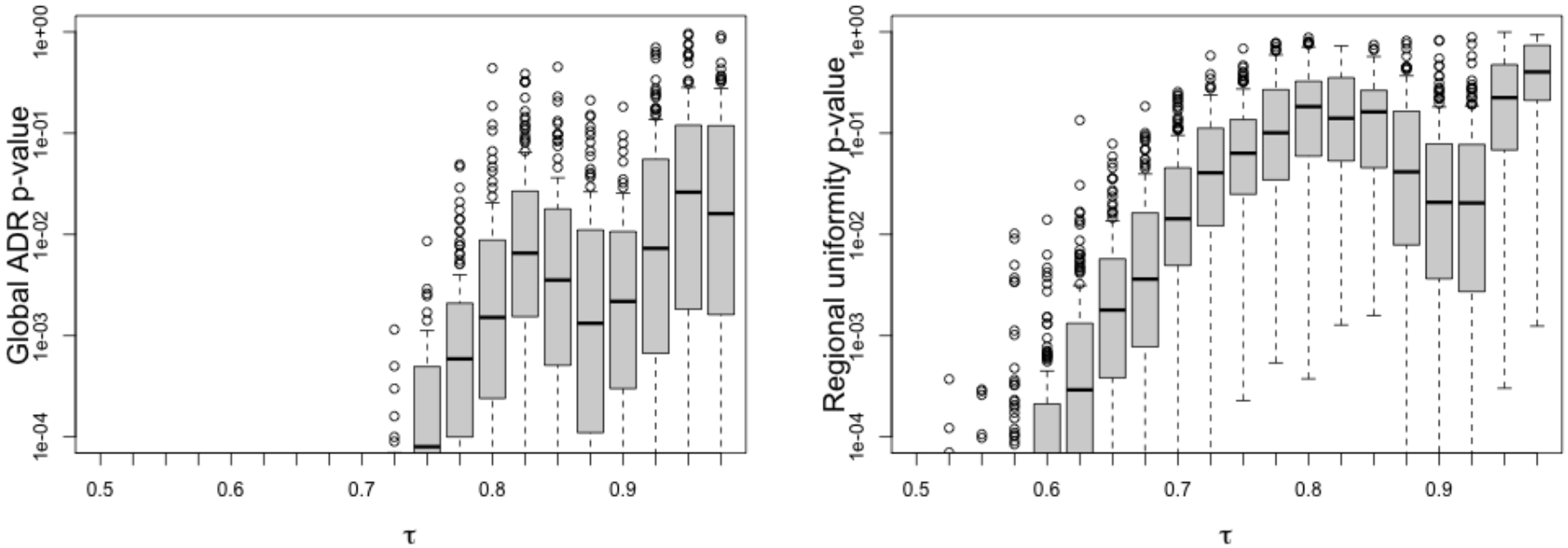


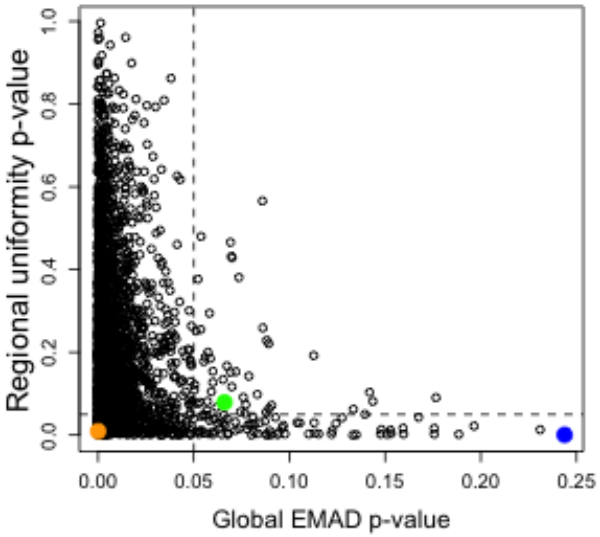
Figure: Box-plots of the global ADR p-values (left) and regional uniformity p-values (right), under **5000 estimated candidate models**.

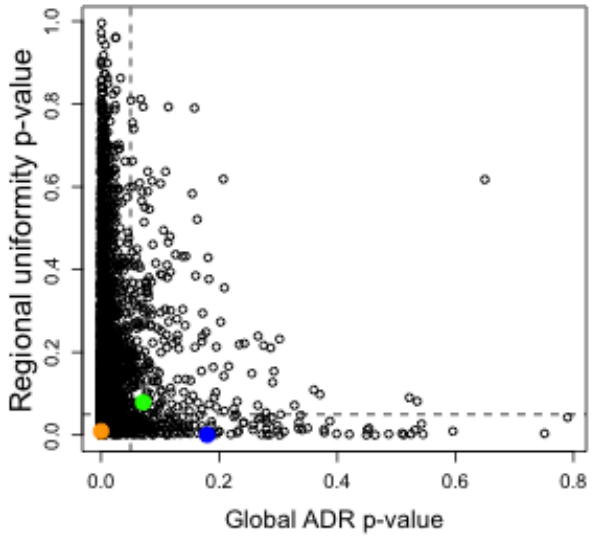
Local summary statistics

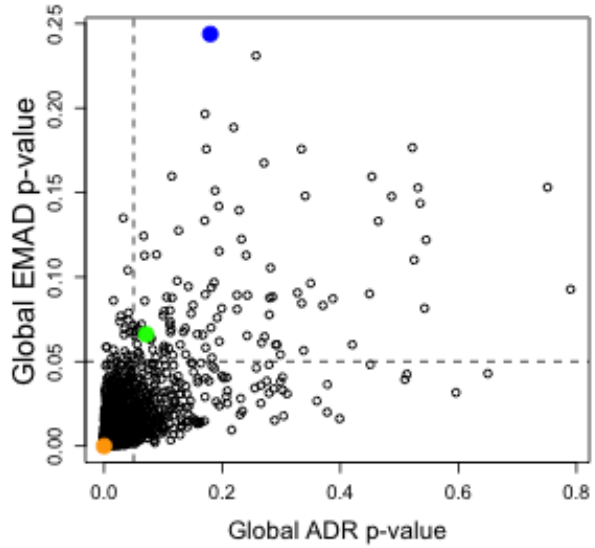
We pick $\tau = 0.825$.

- A further 2500 candidate models are fitted;
- For every model estimate, we evaluate all visual and quantitative goodness-of-fit diagnostics;
- Each bin b has $\approx$ 10–70 **test** points;

Local summary statistics

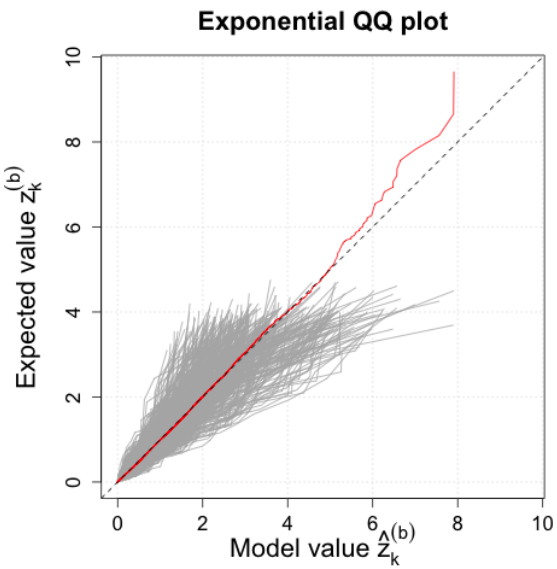


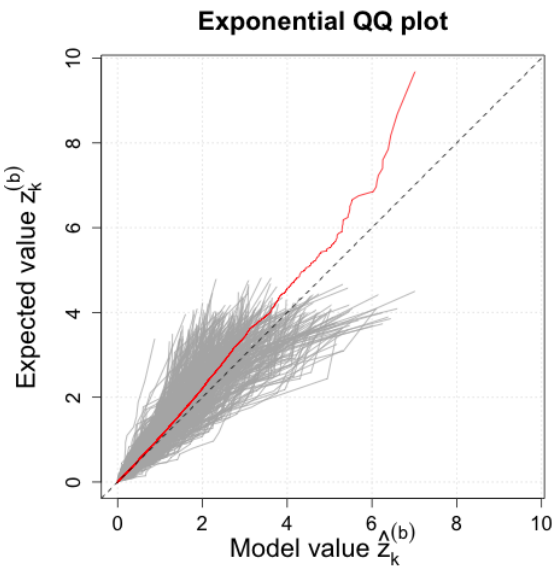


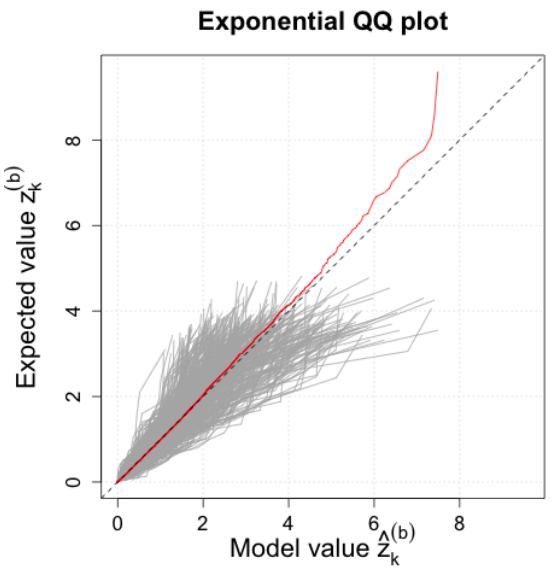


Models 1, 2, and 3.

Pooled exponential QQ plots

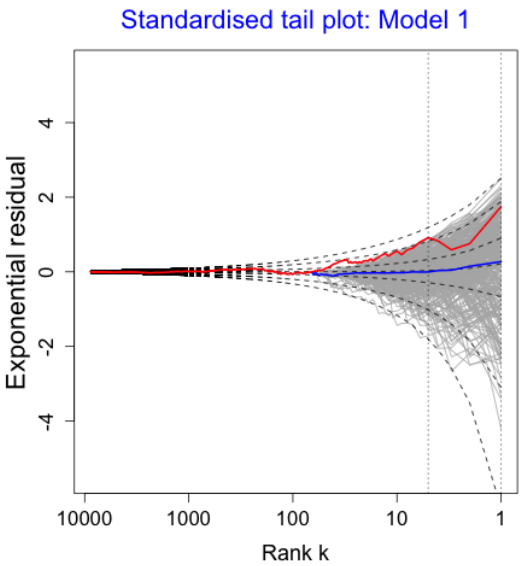


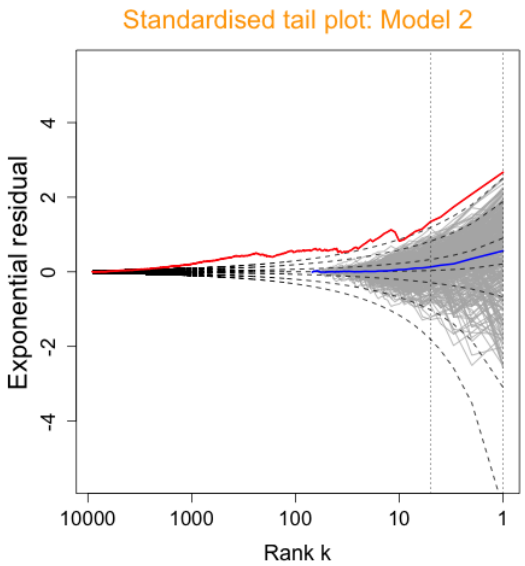


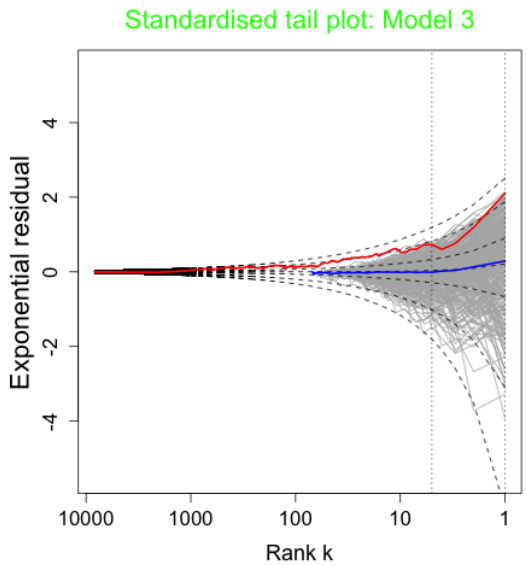


Models 1, 2, and 3.

Standardised tail plots

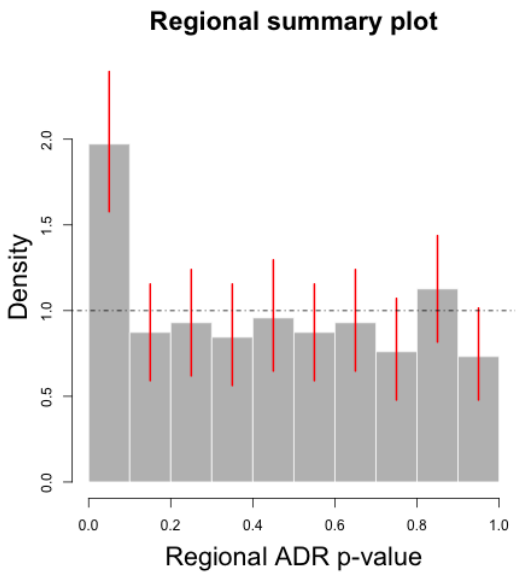


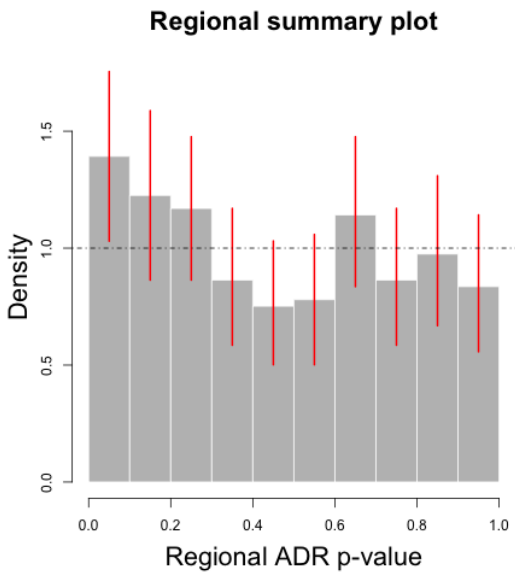


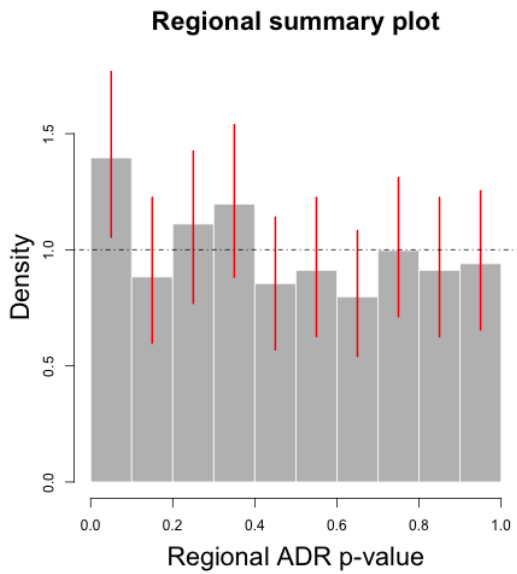


Models 1, 2, and 3.

Regional summary plots

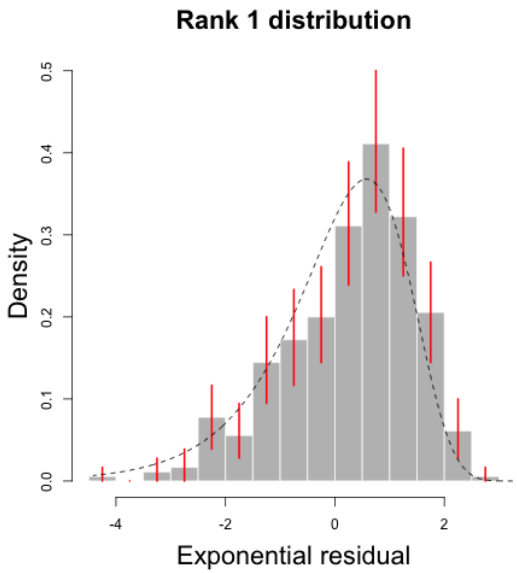


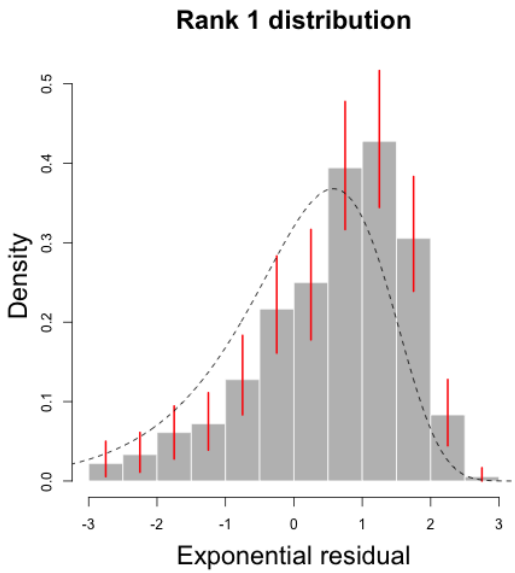


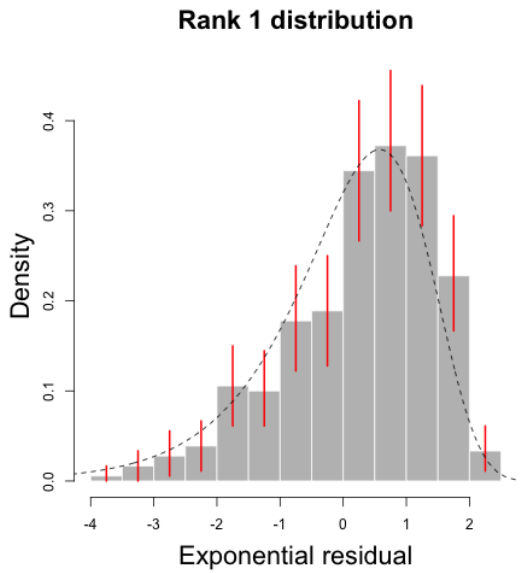


Models 1, 2, and 3.

Rank-wise normalised residual histograms







Models 1, 2, and 3.

Model 3 gives best overall performance.

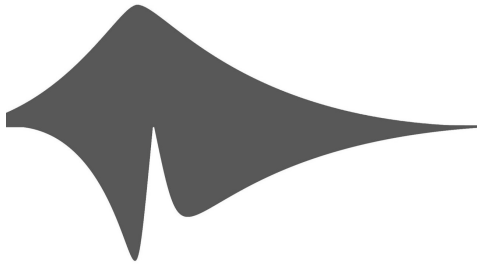
Conclusion

Conclusions:

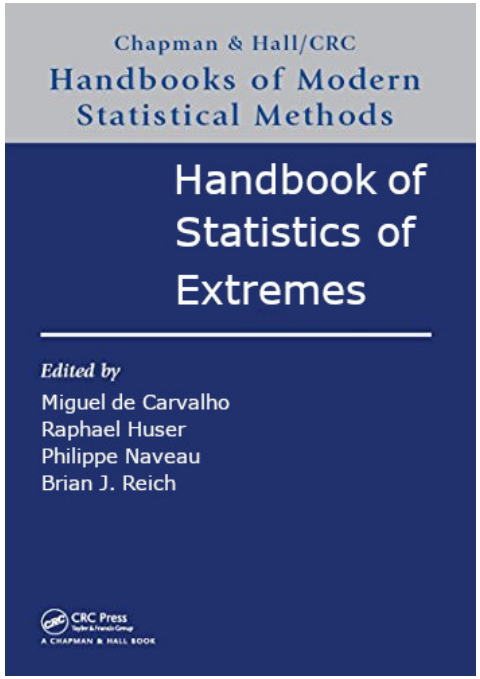
- New diagnostic tools for **extreme value regression models**;
- Not specific to i) extremes, ii) deep neural nets, iii) GPD models;
- Quick consistent model diagnostics, which can help sift through thousands of model fits;
- **Limitation:** Regional diagnostics are **sensitive to binning scheme**, which is application/domain/user specific;
- Paper "Diagnostic Tools for Extreme Value Regression Models" available very soon.

Shameless advertisements

Gimeno-Sotelo, Richards, Hazra, Mhalla, and de Zea Bermudez. (2026).
A Review of Applications of Extreme Value Theory to Environmental Risk Assessment.
Environmental Modelling with Contemporary Statistics: Learning, Directionality, and Space-Time Dynamics, Chapman and Hall/CRC.



Outline



Chapter 21. Richards and Huser (2026a).
Extreme Quantile Regression with Deep Learning.
Code and short courses on GitHub - `https://github.com/Jbrich95/`

Special issue

Home > Collection

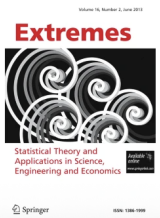
Bridging Heavy Tails and Artificial Intelligence

Participating journal: **Extremes**

Closed for submissions

The increasing frequency and severity of extreme events—such as catastrophic flooding, record-breaking temperatures, and unprecedented heatwaves—has highlighted the urgent need for innovative approaches in risk assessment and modeling. Modern advancements in data-collection techniques have provided increasingly large and complex datasets, which can only be processed used fast and scalable algorithms and computational software. This special issue aims to bridge the gap between Artificial Intelligence (AI) and Extreme Value Theory (EVT) to harness the strengths of both fields and address the growing challenges posed by these extreme events.

Participating journal



Journal

Extremes

Extremes is a dedicated platform for the publication of original research in statistical extreme value theory and its applications across various fields.

Publishing model	**Hybrid**
Journal Impact Factor	**1.1 (2023)**
Downloads	**58k (2024)**
Submission to first decision (median)	**7 days**

https://link.springer.com/collections/haghbdfdhb

References I

Castro-Camilo, D., Huser, R., and Rue, H. (2022). Practical strategies for generalized extreme value-based regression models for extremes. *Environmetrics*, 33(6):e2742.

Cisneros, D., Richards, J., Dahal, A., Lombardo, L., and Huser, R. (2024). Deep graphical regression for jointly moderate and extreme Australian wildfires. *Spatial Statistics*, 59:100811.

Mackay, E. and Jonathan, P. (2023). Modelling multivariate extremes through angular-radial decomposition of the density function. *arXiv preprint arXiv:2310.12711*.

Mackay, E., Murphy-Barltrop, C. J. R., Richards, J., and Jonathan, P. (2025). Deep Learning Joint Extremes of Metocean Variables Using the SPAR Model. *Journal of Offshore Mechanics and Arctic Engineering*, 148(021201).

Majumder, R. and Richards, J. (2026). Semi-parametric bulk and tail regression using spline-based neural networks. *Extremes*, page To appear.

Papastathopoulos, I. and Tawn, J. A. (2013). Extended generalised Pareto models for tail estimation. *Journal of Statistical Planning and Inference*, 143(1):131–143.

Richards, J. and Huser, R. (2026a). Extreme Quantile Regression with Deep Learning. In de Carvalho, M., Huser, R., Naveau, P., and Reich, B. J., editors, *Handbook on Statistics of Extremes*. Chapman & Hall/CRC, Boca Raton, FL.

Richards, J. and Huser, R. (2026b). Regression modelling of spatiotemporal extreme US wildfires via partially-interpretable neural networks. *Journal of Computational and Graphical Statistics*, page To appear.

Fin.

Scan for webpage.