



Extreme value statistics for analysing simulated environmental extremes

Henry Elsom¹ · Matthew Pawley¹

Received: 20 December 2023 / Revised: 30 September 2024 / Accepted: 1 October 2024 /

Published online: 9 November 2024

© The Author(s) 2024

Abstract

We present the methods employed by team ‘Uniofbathtopia’ as part of a competition organised for the 13th International Conference on Extreme Value Analysis (EVA2023), including our winning entry for the third sub-challenge. Our approaches unite ideas from extreme value theory, which provides a statistical framework for the estimation of probabilities/return levels associated with rare events, with techniques from unsupervised statistical learning, such as clustering and support identification. The methods are demonstrated on the data provided for the EVA (2023) Conference Data Challenge – environmental data sampled from the fantasy country of ‘Utopia’ – but the underlying assumptions and frameworks should apply in more general settings and applications.

Keywords Bootstrapping · Extreme value analysis · Max-linear model · Sparse projections · Tail pairwise dependence matrix

AMS 2000 Subject Classifications 62G32

1 Introduction

In recent decades, the field of environmental sciences has experienced significant advancements, particularly through the utilisation of sophisticated modelling techniques to better understand extreme events. Extreme value analysis is the branch of statistical modelling that focuses on quantifying the frequency and severity of very rare events. Notably, the Peaks over Threshold (PoT) approach (Davison and Smith 1990; Pickands 1975), has played a pivotal role in enhancing our comprehension of extreme

✉ Henry Elsom
he294@bath.ac.uk

Matthew Pawley
mtp34@bath.ac.uk

¹ Department of Mathematical Sciences, University of Bath, Claverton Down, Bath BA2 7AY, Somerset, UK

environmental phenomena. Within climate science, significant strides have been made in the modelling of a broad spectrum of variables, including temperature (Clarkson et al. 2023), precipitation (Katz 1999), wind speeds (Kunz et al. 2010; Fawcett and Walshaw 2006), as well as other broader environmental topics including hydrology (Towler et al. 2010; Katz et al. 2002) and air pollution (Gouldsborough et al. 2022).

In this paper, we outline the techniques utilised by the team ‘Uniofbathtopia’ for the data challenge organised for the 13th International Conference on Extreme Value Analysis (EVA2023). A full description of the tasks can be found in the editorial (Rohrbeck et al. 2023). We outline our methodologies for each of the four sub-challenges, in which we complement traditional methods from extreme value statistics with other statistical modelling techniques according to the requirements of each task. The challenges involve the estimation of extreme marginal quantiles, marginal exceedance probabilities and joint tail probabilities within the context of an environmental application, designed on the fictitious country of ‘Utopia’. The organisers of the competition simulated the data using known parameters, so that teams’ models could be validated and compared, and in such a way as to mimic the rich, complex behaviour exhibited by real-world processes. Therefore, we expect that the performance of our proposed methods should extend to general settings and applications.

In the univariate tasks we utilise the generalised Pareto distribution (GPD) and use model-based clustering methods including hierarchical models (Hastie et al. 2009) and mixture models (Fraley and Raftery 2002), as well as Markov chain Monte Carlo (MCMC) for parameter estimation (Coles and Powell 1996). We also use bootstrapping methods for confidence interval estimation (Gilleland 2020). For the multivariate problems our approaches are heavily based on the parametric family of max-linear combinations of regularly varying random variables (Fougères et al. 2013). We introduce novel methods for performing inference for these models, advancing existing approaches (Cooley and Thibaud 2019; Kiriliouk and Zhou 2022) using modern statistical learning techniques including sparsity-inducing projections and clustering. The novel aspects of our work are: exploring MCMC parameter estimation bias for systems with large uncertainty in tail behaviour, and proposing a new estimator for the noise coefficient matrix of a max-linear model based on sparse projections onto the simplex.

The format of the paper is as follows: Section 2 describes our solutions for the univariate challenges with each challenge split into methodology and results sections. Section 3 introduces the requisite background theory from multivariate extremes before outlining the methodological frameworks and the results attained for Challenges 3 and 4. We conclude with some final discussion of our performance in Section 4.

2 Univariate challenges

The first two challenges both involve estimating univariate extreme marginal quantiles, so we initially describe some of the theory that will be used across both tasks. Suppose that a generalised Pareto distribution with scale and shape parameters, σ and ξ respectively, is a suitable model for exceedances of a high threshold u by a variable

Y . Then, for $y > u$,

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

where $\xi, u \in \mathbb{R}, \sigma \in \mathbb{R}^+$ and $(\cdot)_+ = \max(\cdot, 0)$. Given the probability $\zeta_u = \mathbb{P}(Y > u)$, the probability of $Y > y$ can be expressed as

$$\mathbb{P}(Y > y) = \zeta_u \mathbb{P}(Y > y \mid Y > u). \quad (1)$$

We can find the p th percentile of the distribution of Y by rearranging the distribution function to form the quantile function,

$$q(p) = \begin{cases} u + \frac{\sigma}{\xi} \left(\left(\frac{\zeta_u}{1-p} \right)^\xi - 1 \right), & \xi \neq 0, \\ u + \sigma \log \left(\frac{\zeta_u}{1-p} \right), & \xi = 0, \end{cases}$$

where $p > 1 - \zeta_u$ to ensure that $q(p) > u$. However, we can also write down this expression for a return level, that is, the level y_T that is exceeded on average once every T years,

$$y_T = \begin{cases} u + \frac{\sigma}{\xi} ((Tn_{\text{yr}}\zeta_u)^\xi - 1), & \xi \neq 0, \\ u + \sigma \log(Tn_{\text{yr}}\zeta_u), & \xi = 0. \end{cases} \quad (2)$$

where n_{yr} is the number of observations per year. For the purposes of the second challenge, $n_{\text{yr}} = 300$ as we are given that a year in Utopia consists of 12 months and 25 days per month, and there is one observation per day.

These expressions imply that, having chosen an appropriate threshold u , we can estimate quantiles and return levels once we obtain estimates of σ and ξ . The method of obtaining these values is different for different tasks and will be described, for the univariate challenges, in Sections 2.2 and 2.3. We also require an estimate of ζ_u , the probability of an individual observation exceeding the threshold u . We can achieve this by using the empirical probability of Y exceeding u ,

$$\hat{\zeta}_u = \sum_{i=1}^n \mathbb{1}\{y_i > u\}/n, \quad (3)$$

that is, the proportion of the total number of observations that exceed the threshold for observations y_1, \dots, y_n , where for the univariate challenges $n = 21,000$, and where $\mathbb{1}\{y_i > u\}$ is the indicator function,

$$\mathbb{1}\{y_i > u\} = \begin{cases} 1 & \text{if } y_i > u, \\ 0 & \text{otherwise.} \end{cases}$$

2.1 Data

We are interested in the extreme values of the environmental variable Y which, for the two univariate tasks, we denote Y_i for day i . For each day, we also have a vector of covariates $\mathbf{X}_i = (V_{1,i}, \dots, V_{8,i})$ with variables (V_1, \dots, V_4) representing unnamed covariates and (V_5, V_6, V_7, V_8) representing season, wind speed, wind direction and atmosphere respectively. Season is a factor variable, where $V_5 \in \{1, 2\}$, and each season covers half of the year. We perform our methodology for this challenge on each season individually. Given the covariates $\mathbf{X}_1, \dots, \mathbf{X}_n$, Y_1, \dots, Y_n are independent (Rohrbeck et al. 2023). For the first challenge, the data is divided into a training set comprising 70 years' worth of daily environmental data and a test set featuring 100 days of environmental data, showcasing diverse combinations of the covariates. We denote the test data points by $\tilde{\mathbf{x}}_1, \dots, \tilde{\mathbf{x}}_{100}$. A comprehensive description of the dataset for this challenge is available in the competition editorial (Rohrbeck et al. 2023).

2.2 Challenge 1

2.2.1 Methodology

In the first task, we are required to provide point estimates and central 50% confidence intervals for extreme quantiles. As such, we need a model for the distribution of $Y | \mathbf{X}$ in order to estimate the conditional quantiles, q_1, \dots, q_{100} , where

$$\mathbb{P}(Y \leq q_k) | \mathbf{X} = \tilde{\mathbf{x}}_k = 0.9999.$$

We approach this problem with a two step strategy. We first partition the covariate space into clusters and then analyse the extremes of each cluster individually. In each resulting cluster, it is assumed that the extreme values can be adequately modelled using a GPD, characterised by common scale and shape parameters, along with a common threshold and exceedance probability. To implement this, we propose partitioning the observations into groups by fitting a Gaussian mixture model to the environmental covariates. The dataset contained some missing values. We assumed that the missing observations were missing completely at random and removed them from the analysis, which was a valid assumption based on the editorial (Rohrbeck et al. 2023).

We perform the clustering of the covariates for each season individually by using the method in Fraley and Raftery (2003). Let $\mathbf{x}_1^{(s)}, \dots, \mathbf{x}_N^{(s)}$, where $N = n/2$, denote the observations for season $s \in \{1, 2\}$. We then assume that the covariates are sampled from a mixture of multivariate normal distributions. More formally, we want to estimate a d -dimensional Gaussian mixture model (where $d = 7$ in this challenge) with $J^{(s)}$ components, mean vectors $\boldsymbol{\mu}_1^{(s)}, \dots, \boldsymbol{\mu}_{J^{(s)}}^{(s)} \in \mathbb{R}^d$, mixture probabilities $\boldsymbol{\alpha}^{(s)} = \{\alpha_1^{(s)}, \dots, \alpha_{J^{(s)}}^{(s)}\}$, where $\alpha_j^{(s)} > 0$ and $\sum_{j=1}^{J^{(s)}} \alpha_j^{(s)} = 1$, and covariance matrices $\boldsymbol{\Sigma}_1^{(s)}, \dots, \boldsymbol{\Sigma}_{J^{(s)}}^{(s)} \in \mathbb{R}^{d \times d}$, which are symmetric and positive definite. In the remainder of the subsection, we drop the season index s for brevity. The probability density

function of the observation \mathbf{x}_i can then be written as

$$p(\mathbf{x}_i \mid \Psi) = \sum_{j=1}^J \alpha_j \phi(\mathbf{x}_i; \boldsymbol{\mu}_j, \boldsymbol{\Sigma}_j) \quad (4)$$

where $\Psi = \{\boldsymbol{\alpha}, \boldsymbol{\mu}, \boldsymbol{\Sigma}\}$ are the parameters of the mixture model and we use $\phi(\cdot; \boldsymbol{\mu}_j, \boldsymbol{\Sigma}_j)$ to denote the probability density function of the Gaussian distribution with mean vector $\boldsymbol{\mu}_j$ and covariance matrix $\boldsymbol{\Sigma}_j$. Assuming that an appropriate fixed J has been found for each season, the mixture model parameters Ψ are unknown and to be found. In order to achieve this, we use the expectation maximisation (EM) algorithm to perform maximum likelihood estimation (McLachlan and Peel 2000), which is executed using the `mclust` package in R (Fraley and Raftery 2003). The optimum number of clusters is then found by using the elbow method; we derive the negative log-likelihood for each choice of J , and identify the point at which this value begins to plateau.

In addition to the cluster parameters, the EM algorithm gives a cluster allocation for the observations. We use this allocation to split the observations of Y , and we denote by $\mathcal{Y}^{(j)}$ the data points in the j th cluster ($j = 1, \dots, J$). We are then able to perform extreme value analysis on each set of points $\mathcal{Y}^{(j)}$ separately, by fitting a GPD, with parameters $\hat{\sigma}_j$ and $\hat{\xi}_j$, to the extremal data, using maximum likelihood estimation, as outlined at the start of Section 2. We make the assumption that the effects of the covariates on Y are entirely captured by the cluster assignments. The threshold, \hat{u}_j , is chosen by interpreting mean residual life plots and parameter stability plots (Coles 2001).

The cluster estimates are then used to estimate q_k . We introduce a latent variable Z_k , where $Z_k = j$ refers to $\tilde{\mathbf{x}}_k$ being allocated to the j th cluster. Using the law of total probability, we then write $\mathbb{P}(Y > q_k \mid X = \tilde{\mathbf{x}}_k)$ as

$$\mathbb{P}(Y > q_k \mid X = \tilde{\mathbf{x}}_k) = \sum_{j=1}^J \mathbb{P}(Z_k = j \mid X = \tilde{\mathbf{x}}_k) \mathbb{P}(Y > q_k \mid X = \tilde{\mathbf{x}}_k, Z_k = j). \quad (5)$$

We can write the first probability in this expression using Eq. 4,

$$\mathbb{P}(Z_k = j \mid X = \tilde{\mathbf{x}}_k) = \frac{\alpha_j \phi(\tilde{\mathbf{x}}_k; \boldsymbol{\mu}_j, \boldsymbol{\Sigma}_j)}{p(\tilde{\mathbf{x}}_k \mid \Psi)},$$

and using Eq. 1, it is possible to express the second probability as

$$\mathbb{P}(Y > q_k \mid Z_k = j, X = \tilde{\mathbf{x}}_k) = \mathbb{P}(Y > q_k \mid Y > \hat{u}_j, Z_k = j) \mathbb{P}(Y > \hat{u}_j \mid Z_k = j).$$

The exceedance probability, $\mathbb{P}(Y > \hat{u}_j \mid Z_k = j)$ is calculated empirically using Eq. 3, and the other probability in the expression is found by fitting a GPD to the clusters. We then have all the components in Eq. 5 to find q_k that satisfies $\mathbb{P}(Y > q_k \mid X = \tilde{\mathbf{x}}_k) = 0.0001$.

The challenge also requires central 50% confidence intervals for the estimates of these extreme conditional quantiles. For each set of points in a cluster, $\mathcal{Y}^{(j)}$, we are able to find an estimate of the quantile by performing a jackknife resampling (Efron and Stein 1981) of the extremal data and use this to fit the GPD. In the jackknife resampling method, sets of samples are created by leaving out one observation at a time and calculating the quantile point estimate based on the remaining observations. For each observation i in the extremal dataset $\mathcal{Y}^{(j)} \mid \mathcal{Y}^{(j)} > \hat{u}_j$, we create a new dataset $\mathcal{Y}_{(-i)}^{(j)} \mid \mathcal{Y}^{(j)} > \hat{u}_j$ by excluding the i th observation. We then fit a new GPD to this data and, using these parameter estimates, calculate the quantile point estimate in Eq. 2. Once we have undertaken this process for every observation in the dataset, we calculate the empirical 25th and 75th percentiles for the jackknife samples.

2.2.2 Results

As mentioned above, we perform clustering separately for each of the two seasons - Season 1 and Season 2. Elbow plots are used to determine suitable number of clusters; these results are displayed in Fig. 1. It is clear from these figures that the negative log-likelihood does not obviously level off for up to 10 clusters. After balancing the trade-off between maximising the number of covariate clusters, and maximising the number of data points within each cluster, the number of covariate clusters was determined to be $J = 5$ for each season. By choosing $J = 5$, the average number of observations in each cluster was 2,580 (with a standard deviation of 535).

The performance of this method was disappointing (see editorial Rohrbeck et al. 2023), and there are a few reasons why this could have been the case. Firstly, there were no clearly defined clusters. This points to the fact that the fitted Gaussian mixture distribution may have been unable to fully capture the non-Gaussian distribution of the covariate values. In Fig. 1, the point at which the negative log-likelihood plateaus is not clear, illustrating this argument. The second reason could be that covariates were not used when fitting the GPDs to the extreme data in each cluster, only for cluster assignment. A GPD with covariate models for each cluster may have resulted

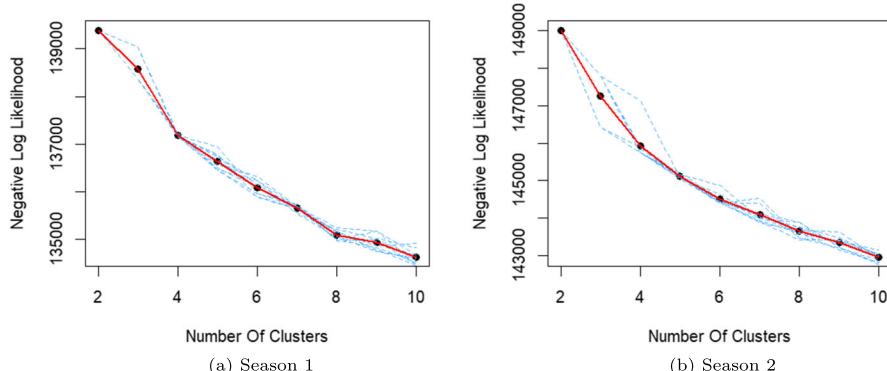


Fig. 1 An elbow plot to determine an optimum number of clusters for each season. The figure displays the negative log-likelihood for different simulations (blue) as well as the average (red)

in better quantile estimates. A further development could be to explore generalised Pareto regression trees, as they are able to perform clustering using trends in the covariates (Farkas et al. 2021).

2.3 Challenge 2

2.3.1 Methodology

In the second challenge, we are required to estimate the marginal quantile $q \in \mathbb{R}$ such that

$$\mathbb{P}(Y > q) = \frac{1}{300T},$$

where $T = 200$, that is, an estimate of the 200 year return level for Amaurot. This challenge was assessed using a loss function that penalises under-estimating the quantile more than over-estimating it. Formally, for a given estimate \hat{q} and the true marginal quantile q , the loss is calculated as,

$$L(q, \hat{q}) = \begin{cases} 0.9(0.99q - \hat{q}) & \text{if } 0.99q > \hat{q}, \\ 0, & \text{if } |q - \hat{q}| \leq 0.01q, \\ 0.1(\hat{q} - 1.01q), & \text{if } 1.01q < \hat{q}. \end{cases} \quad (6)$$

This replicates real-world scenarios. For example, in a hydrology context, over-estimating flood defences leads to increased costs, while under-estimating them results in severe consequences, such as fatalities. After completing the challenge, we were given the correct quantile $q = 196.6$ and so we can substitute this value into the loss function, as displayed in Fig. 2.

For this analysis, we make the underlying assumption that the extreme values of Y are independent and identically distributed (IID), that is, we ignore any dependence on X . This assumption is made to simplify the modelling process. We then assume that the exceedances of Y above a high threshold u follow a GPD,

$$Y - u \mid Y > u \sim \text{GPD}(\sigma, \xi).$$

Using Eq. 2, the 200-year return level is then

$$y_{200} = \hat{u} + \frac{\hat{\sigma}}{\hat{\xi}}((200 \times 300\hat{\zeta}_u)^{\hat{\xi}} - 1),$$

where we have substituted $T = 200$ and $n_y = 300$ as we have 300 daily observations a year.

To define a Bayesian framework, we choose prior distributions for σ and ξ , $\sigma \sim \text{Gamma}(4, 1)$, $\xi \sim \text{Normal}(0, 1)$. These prior distributions are selected to reflect minimal prior knowledge about the exact parameter values (whilst ensuring that σ remain positive). Subsequent analysis demonstrated that the parameter estimates were

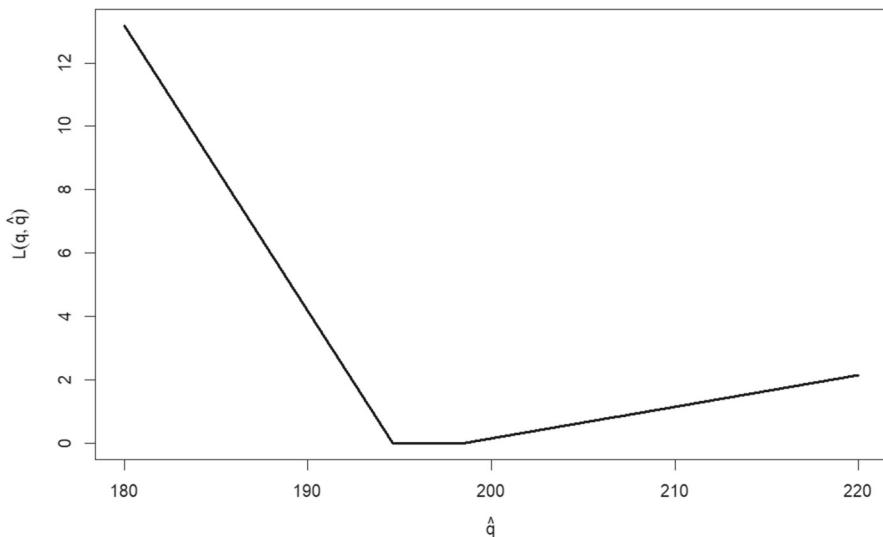


Fig. 2 A diagram of the loss function using the true quantile value $q = 196.6$

robust and largely unaffected by the specific choice of priors, which indicated that the data provided sufficient information to primarily determine the posterior distributions.

We use Markov chain Monte Carlo methods (Coles and Powell 1996; Gelfand and Smith 1990) to sample from the resulting posterior distribution of (σ, ξ) . This method was preferred over maximum likelihood estimation as we wanted a full distribution of the parameters, allowing for more insight into the uncertainty of the estimates.

To assess how well our approach recovers the true parameter values, we generate sets of simulated data from GPDs with parameters that were representative of the ones sampled via our MCMC algorithm. Formally, for each simulated dataset, we sample values $\tilde{\sigma}$ and $\tilde{\xi}$, and then generate data points $\tilde{w}_1, \tilde{w}_2, \dots, \tilde{w}_m$ from a $\text{GPD}(\tilde{\sigma}, \tilde{\xi})$, where m is equal to the number of exceedances of Y above u . We considered two approaches to sample $\tilde{\sigma}$ and $\tilde{\xi}$, and we generated 1000 datasets for each. The first approach is to sample these parameters from the posterior distribution. The other approach is to sample the parameters from uniform distributions, $\tilde{\sigma} \sim \text{Unif}(11, 18)$ and $\tilde{\xi} \sim \text{Unif}(-0.15, 0.20)$; these limits were selected based on the posterior samples. For each generated dataset, we again run our MCMC algorithm and store the posterior mean estimate for $\tilde{\sigma}$ and $\tilde{\xi}$. Finally, we apply the mean corrections to the initial GPD parameter estimates to calculate an adjusted return level.

2.3.2 Results

We determine a threshold of $\hat{u} = 110$ by using mean excess plots, resulting in $m = 180$. We can then use this to calculate the empirical estimate for the exceedance probability $\hat{\xi}_u = 180/21000 \approx 0.0086$. Figure 3 shows the posterior samples of the shape and scale distributions for the initial GPD fit. The range of these distributions are then used to inform the sampling boundaries for the simulated datasets as described

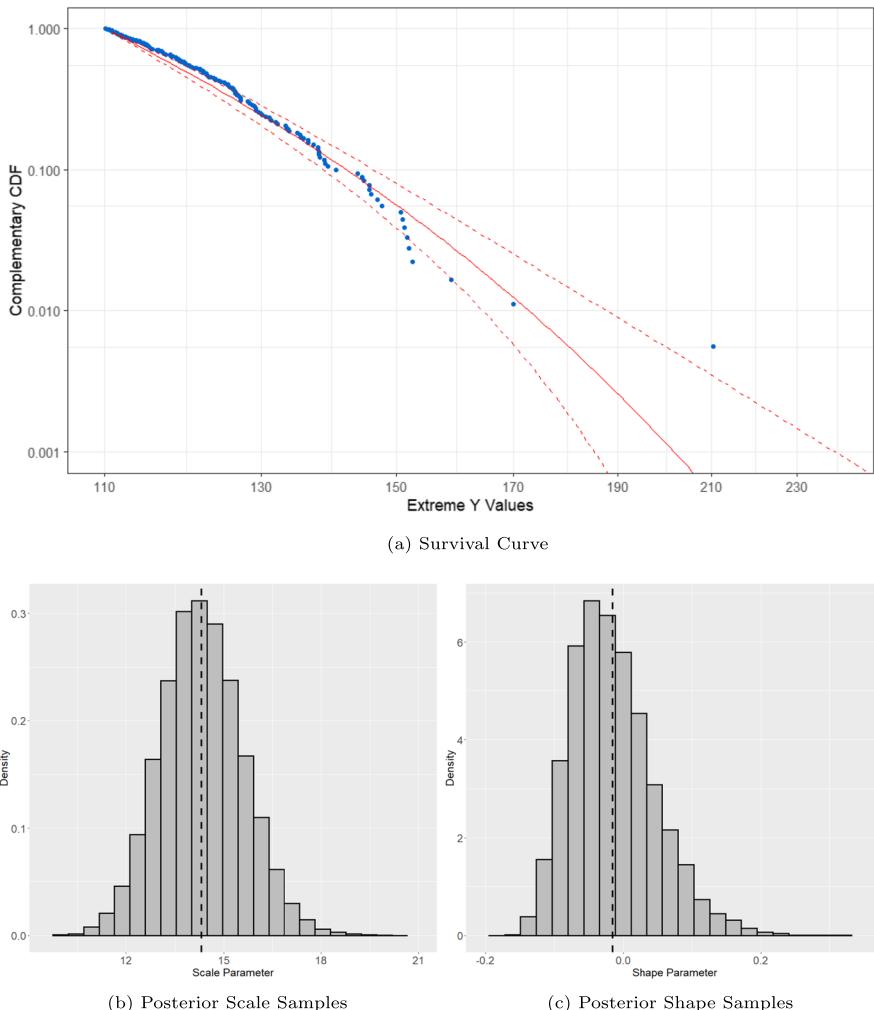


Fig. 3 A GPD fitted to extreme Y values above threshold 110, with a) a survival curve fitted using the mean of the posterior parameter values, with a 90% confidence interval as dashed lines, and b), c) the respective posterior distributions of sampled scale and shape parameters, with the mean values as dashed lines

in Section 2.3.1. We also plot a survival function to assess the model fit (Fig. 3), where the dashed red lines reflect a 90% credible interval for the fit. We find that the posterior distribution of the shape parameter spanned both negative and positive values, indicating uncertainty in the tail behaviour of the underlying distribution. It was also clear from inspecting smaller values of the extremes in the survival curve, that the GPD fit could be improved.

We opted for our second approach to create the simulated datasets, that is, we sampled the parameters from uniform distributions to make the corrections for our final results. We choose this approach as, by employing uniform coverage of the

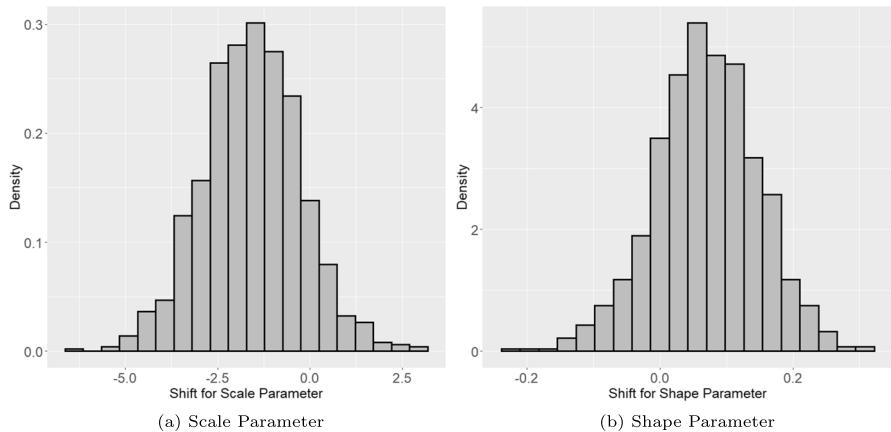


Fig. 4 A histogram of residuals of predicted and simulated scale (left) and shape (right) parameters. These residuals are then used as a bias correction for the return level estimate

parameter space, we return a more risk-averse return level with respect to the challenge loss function. The distributions of the residuals of the mean predicted parameters and the true (simulated) parameters are displayed in Fig. 4. There are significant negative residuals for the scale parameter and significant positive residuals for the shape parameter. We wanted to use these residual distributions to make a correction to the initial parameter estimates. We apply the mean corrections to the initial posterior parameter values and compared the mean return levels. These are shown in Fig. 5.

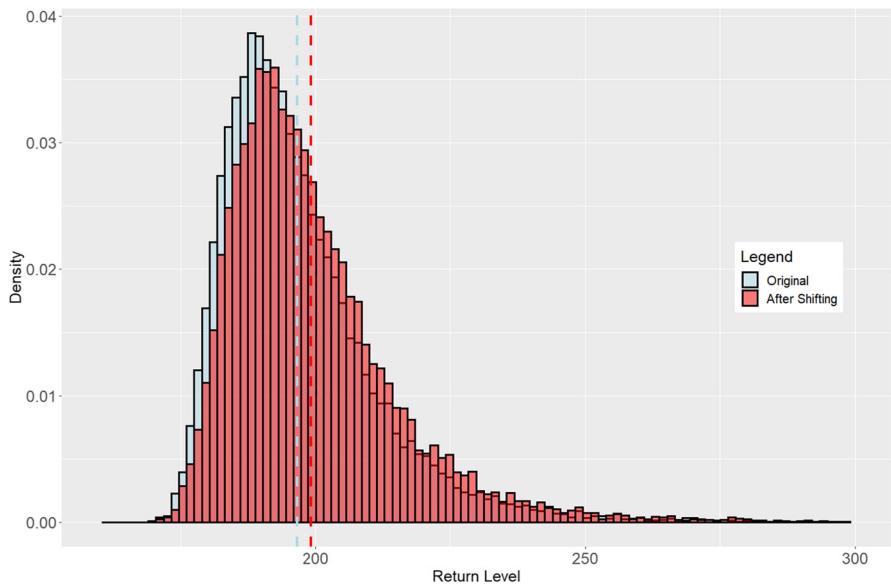


Fig. 5 The return level samples before shifting (blue) and after shifting (red). The dashed lines denote the mean of the samples

This non-standard approach was used for simplicity, however, some further work could be to more carefully incorporate these results into the initial posterior parameter distributions. This gives an overall increase in the average return level from 196.4 to 199.4.

The correct quantile value was $q = 196.6$. We were interested to find that the initial GPD fit to the data gave a return level of 196.4, which led to a loss of $L(196.6, 196.4) = 0$. It is possible that small effects of the covariates are compensated for by other factors, giving a correct return level under the misspecified model assumption. Although our final answer, with the correction, increased our return level, it also increased the loss from $L(196.6, 196.4) = 0$ to $L(196.6, 199.4) = 0.0834$ (using Eq. 6). Before the correct quantile value was announced, we were encouraged that the method resulted in an increase in the return level, as an over-estimate was preferable to an under-estimate in the context of the competition. It is worth noting that if we had opted to sample the parameters of the simulated data from the posterior distributions, we would have got an average return level of 198.7, which would have given a smaller loss, $L(196.6, 198.7) = 0.0134$. The simplicity of the model, given the underlying assumption about the data, creates ambiguity regarding whether the large residuals in the parameter estimates are caused by bias or by model misspecification. For future work, it would be appropriate to incorporate a more sophisticated model of the covariates to reduce model misspecification and to investigate whether any residuals in parameter estimates are due to bias.

3 Multivariate challenges

In Challenges 3 and 4 we are presented with a d -dimensional random vector $X = (X_1, \dots, X_d) \sim F_X$, representing the value of an environmental variable at d sites in Utopia, where F_X is an unknown joint distribution function. Our goal is to estimate the probability that X lies in a given extreme ‘failure region’ based on a sample of independent observations of X . The failure regions of interest are such that certain components of X are simultaneously large while all remaining components (if any) are of lower order. The inherent difficulty of the task stems from the fact that the events’ return periods are of similar order or even significantly longer than the observation period over which the data are collected; empirical methods based solely on the relative frequency of event occurrences are fruitless. Instead, we use the observed data to infer an estimate for the probabilistic structure of the joint tail of X and subsequently compute tail event probabilities under this model. This encapsulates the philosophy of multivariate extreme value statistics.

In the absence of prior knowledge about the physical processes driving the environment of Utopia, we are compelled to recourse to data-driven, statistical learning methods for multivariate extremes. The particular tools we choose will depend on the nature and difficulties of the task at hand. In Challenge 3, the failure regions are defined by different subsets of variables being large, thereby placing emphasis on accurately modelling the so-called extremal directions of X . The salient characteristic of Challenge 4 is its high dimensionality, which calls for the utilisation of dimension

reduction techniques, such as clustering, in order to overcome the curse of dimensionality inherent to tail dependence estimation.

3.1 Background

3.1.1 Multivariate regular variation and the angular measure

Let X denote a random vector that takes values in the positive orthant $\mathbb{R}_+^d := [0, \infty)^d$. As is commonly done in multivariate extremes, we work in the framework of multivariate regular variation (MRV).

Definition 1 We say that X is *multivariate regularly varying* with index $\alpha > 0$, denoted $X \in \text{RV}_+^d(\alpha)$, if it satisfies the following (equivalent) definitions (Resnick 2007, Theorem 6.1):

1. There exists a sequence $b_n \rightarrow \infty$ and a non-negative Radon measure ν_X on $\mathbb{E}_0 := [0, \infty]^d \setminus \{\mathbf{0}\}$ such that

$$n\mathbb{P}(b_n^{-1}X \in \cdot) \xrightarrow{\text{v}} \nu_X(\cdot), \quad (n \rightarrow \infty), \quad (7)$$

where $\xrightarrow{\text{v}}$ denotes vague convergence in the space of non-negative Radon measures on \mathbb{E}_0 . The *exponent measure* ν_X is homogeneous of order $-\alpha$, i.e. $\nu_X(s \cdot) = s^{-\alpha} \nu_X(\cdot)$ for any $s > 0$.

2. For any norm $\|\cdot\|$ on \mathbb{R}^d , there exists a sequence $b_n \rightarrow \infty$ and a finite *angular measure* H_X on $\mathbb{S}_+^{d-1} := \{\mathbf{x} \in \mathbb{R}_+^d : \|\mathbf{x}\| = 1\}$ such that for $(R, \Theta) := (\|X\|, X/\|X\|)$,

$$n\mathbb{P}((b_n^{-1}R, \Theta) \in \cdot) \xrightarrow{\text{v}} \nu_\alpha \times H_X(\cdot), \quad (n \rightarrow \infty), \quad (8)$$

in the space of non-negative Radon measures on $(0, \infty] \times \mathbb{S}_+^{d-1}$, where $\nu_\alpha((x, \infty)) = x^{-\alpha}$ for any $x > 0$.

The limit measures ν_X and H_X are related via

$$\begin{aligned} \nu_X(\{\mathbf{x} \in \mathbb{E}_0 : \|\mathbf{x}\| > s, \mathbf{x}/\|\mathbf{x}\| \in \cdot\}) &= s^{-\alpha} H_X(\cdot), \\ \nu_X(d\mathbf{r} \times d\theta) &= \alpha r^{-\alpha-1} dr dH_X(\theta). \end{aligned}$$

The probabilistic tail of X decomposes into a univariate α -regularly varying radial component (Resnick 2007, Theorem 3.6) that is asymptotically independent of the angular component. The angular measure represents the limiting distribution of the angular component and encodes all information about the tail dependence structure.

The MRV property implies that the margins of X are heavy-tailed with a common tail index. Henceforth, assume that the components of $X \in \text{RV}_+^d(\alpha)$ are Fréchet distributed with shape parameter α , that is $\mathbb{P}(X_i < x) = \Psi_\alpha(x) := \exp(-x^{-\alpha})$ for $x > 0$ and $i = 1, \dots, d$. (The data for Challenges 3 and 4 are on Gumbel margins but can be accommodated into our framework following a suitable transformation.) Moreover, we choose $\|\cdot\| = \|\cdot\|_\alpha$, the L_α -norm on \mathbb{R}^d , and specify that the normalising

sequence in Eq. 8 is $b_n = n^{1/\alpha}$. With these particular choices the marginal variables have unit scale (Klüppelberg and Krali 2021, Definition 4) and $H_X(\mathbb{S}_+^{d-1}) = d$.

The problem of modelling the angular measure has attracted considerable attention in recent years – a survey of the related literature can be found in Engelke and Ivanovs (2021). One research avenue concerns learning which sets of variables may be concurrently extreme; this can be posed as a support detection problem (Goix et al. 2017; Simpson et al. 2020). Consider the index set $\mathbb{V}(d) := \{1, \dots, d\}$ and denote by $\mathcal{P}_d^* = \mathcal{P}(\mathbb{V}(d)) \setminus \emptyset$ its power set excluding the empty set. A set $\beta \in \mathcal{P}_d^*$ is termed an extremal direction of X if H_X places mass on the subspace

$$C_\beta = \{\mathbf{w} \in \mathbb{S}_+^{d-1} : w_i > 0 \iff i \in \beta\} \subseteq \mathbb{S}_+^{d-1}.$$

Another branch of research aims at developing dimension reduction techniques for analysing the angular measure in high dimensions. To this end, one often considers a summary of the full dependence structure encoded in a matrix of pairwise extremal dependence metrics. One such matrix, originally proposed in Larsson and Resnick (2012) and later popularised by Cooley and Thibaud (2019), is the tail pairwise dependence matrix (TPDM). The TPDM of $\mathbf{X} \in \text{RV}_+^d(2)$ is the $d \times d$ matrix $\Sigma = (\sigma_{ij})$ with entries

$$\sigma_{ij} = \int_{\mathbb{S}_+^{d-1}} \theta_i \theta_j \, dH_X(\boldsymbol{\theta}), \quad (i, j = 1, \dots, d). \quad (9)$$

The diagonal entries are the squared marginal scales, i.e. $\sigma_{ii} = 1$ for $i = 1, \dots, d$. The off-diagonal entries measure extremal dependence between the associated pairs of variables. In particular, $\sigma_{ij} = 0$ if and only if X_i and X_j are asymptotically independent. For asymptotically dependent variables the magnitude of σ_{ij} represents the dependence strength. An important property of the TPDM is its complete positivity (Cooley and Thibaud 2019, Proposition 5).

Definition 2 A matrix S is *completely positive* (CP) if there exists a matrix A with non-negative entries such that $S = AA^T$. We call AA^T a *CP-decomposition* of S and A a *CP-factor* of S .

This property connects the TPDM to the model class introduced in the following section.

3.1.2 The max-linear model for multivariate extremes

Our proposed methods for Challenges 3 and 4 employ the max-linear model, a parametric model based on the class of random vectors constructed by max-linear combinations of independent Fréchet random variables (Fougères et al. 2013). This model is appealing for several reasons. First, it is flexible, in the sense that any regularly varying random vector can be arbitrarily well-approximated by a max-linear model with sufficiently many parameters (Fougères et al. 2013). Since neither Challenge 3 nor Challenge 4 provide any prior information about the underlying data-generating processes, it is preferable to avoid imposing overly restrictive assumptions on the tail dependence structure. Secondly, although the number of parameters grows rapidly –

at least $\mathcal{O}(d)$ but often even $\mathcal{O}(d^2)$ – efficient inference procedures are available even in high dimensions. Scalability is critical for Challenge 4. Finally, extremal directions and failure probabilities can be straightforwardly identified and computed directly from the model parameters (Kiriliouk and Zhou 2022).

Definition 3 For some $q \geq 1$ and $\alpha > 0$, let $\mathbf{Z} = (Z_1, \dots, Z_q)$ be a random vector with independent components $Z_1, \dots, Z_q \sim \Psi_\alpha$ and let $A = (a_{ij}) \in \mathbb{R}_+^{d \times q}$ be a deterministic matrix. If

$$X_i := \sum_{j=1}^q a_{ij} Z_j, \quad (i = 1, \dots, d),$$

then $\mathbf{X} = (X_1, \dots, X_d)$ is said to be *max-linear* with *noise coefficient matrix* A , denoted $\mathbf{X} \sim \text{MaxLinear}(A; \alpha)$, and we write $\mathbf{X} = A \circ \mathbf{Z}$.

Cooley and Thibaud (2019) show that $\mathbf{X} = A \circ \mathbf{Z} \in \text{RV}_+^d(\alpha)$ and

$$H_{\mathbf{X}}(\cdot) = \sum_{j=1}^q \|a_j\|_\alpha^\alpha \delta_{a_j/\|a_j\|_\alpha}(\cdot), \quad (10)$$

where $\delta_x(A) := \mathbb{1}\{\mathbf{x} \in A\}$ is the Dirac mass function. The angles along which extremes can occur are (in the limit) precisely the q self-normalised columns of A . Therefore $\beta \in \mathcal{P}_d^*$ is an extremal direction of \mathbf{X} if and only if there exists $j \in \{1, \dots, q\}$ such that $a_j/\|a_j\|_\alpha \in C_\beta$. When it comes to model fitting, the testing procedure of Kiriliouk (2020) can provide guidance for choosing q ; for our purposes, it either represents a tuning parameter (Challenge 3) or takes a fixed value owing to computational/algorithmic restrictions (Challenge 4). Substituting Eq. 10 into Eq. 9, we observe that $\mathbf{X} \sim \text{MaxLinear}(A, 2)$ has TPDM $\Sigma_{\mathbf{X}} = AA^T$. In other words, the noise coefficient matrix is a CP-factor of the model TPDM. Conversely, given an arbitrary random vector $\mathbf{X} \in \text{RV}_+^d(2)$ with TPDM Σ , any CP-factor A of Σ parametrises a max-linear model with identical pairwise tail dependence metrics to \mathbf{X} .

Kiriliouk and Zhou (2022) give examples of classes of tail events $\mathcal{R} \subset \mathbb{E}_0$ for which $\mathbb{P}(\mathbf{X} \in \mathcal{R})$ can be well-approximated by a function of the parameter matrix A . With a view to Challenges 3 and 4, we focus on tail events where \mathbf{X} is large in the $s \leq d$ components indexed by $\beta = \{\beta_1, \dots, \beta_s\} \in \mathcal{P}_d^*$, while all $d - s$ remaining components are of lower order. Formally, for $\mathbf{u} = (u_1, \dots, u_s) \in \mathbb{R}_+^s$ a vector of high thresholds and $\mathbf{l} \in \mathbb{R}_+^{d-s}$ a vector of comparatively low thresholds, we consider

$$\mathcal{C}_{\beta, \mathbf{u}, \mathbf{l}} := \{\mathbf{x} \in \mathbb{E}_0 : \mathbf{x}_\beta > \mathbf{u}, \mathbf{x}_{-\beta} < \mathbf{l}\}, \quad \mathbb{P}(\mathbf{X} \in \mathcal{C}_{\beta, \mathbf{u}, \mathbf{l}}) = \mathbb{P}(\mathbf{X}_\beta > \mathbf{u}, \mathbf{X}_{-\beta} < \mathbf{l}).$$

When $\beta = \mathbb{V}(d)$ the threshold vector \mathbf{l} is superfluous and is omitted. Kiriliouk and Zhou (2022) specify an approximate formula for $\mathbb{P}(A \circ \mathbf{Z} \in \mathcal{C}_{\mathbb{V}(d), \mathbf{u}})$ in terms of A . We derive an analogous formula for $\mathbb{P}(A \circ \mathbf{Z} \in \mathcal{C}_{\beta, \mathbf{u}, \mathbf{l}})$ for general $\beta \in \mathcal{P}_d^*$ stated as

follows: if $X \sim \text{MaxLinear}(A; \alpha)$, then

$$\mathbb{P}(X \in \mathcal{C}_{\beta, \mathbf{u}, \mathbf{l}}) \approx \hat{\mathbb{P}}(X \in \mathcal{C}_{\beta, \mathbf{u}, \mathbf{l}}) := \sum_{j: \frac{\mathbf{a}_j}{\|\mathbf{a}_j\|_\alpha} \in C_\beta} \min_{i=1, \dots, s} \left(\frac{a_{\beta_i, j}}{u_i} \right)^\alpha. \quad (11)$$

The approximation should be understood as being valid as $u_1, \dots, u_s \rightarrow \infty$ while \mathbf{l} remains fixed. Note that the entries of \mathbf{l} do not appear in the right-hand side of Eq. 11. This reflects the fact that, asymptotically, the magnitude of the probability of the event $\{X_\beta > \mathbf{u}\} \cap \{X_{-\beta} < \mathbf{l}\}$ is predominantly determined by the threshold exceedance event $\{X_\beta > \mathbf{u}\}$ while the relative contribution of the threshold non-exceedance event $\{X_{-\beta} < \mathbf{l}\}$ vanishes. Henceforth, the threshold \mathbf{l} may at times be suppressed for notational convenience. From Eq. 7 we have that, provided u_1, \dots, u_s are sufficiently large,

$$\mathbb{P}(X \in \mathcal{C}_{\beta, \mathbf{u}, \mathbf{l}}) = \frac{1}{n} \left[n \mathbb{P} \left(\frac{X}{n^{1/\alpha}} \in \frac{\mathcal{C}_{\beta, \mathbf{u}, \mathbf{l}}}{n^{1/\alpha}} \right) \right] \approx \frac{1}{n} v_X \left(\frac{\mathcal{C}_{\beta, \mathbf{u}, \mathbf{l}}}{n^{1/\alpha}} \right) = v_X(\mathcal{C}_{\beta, \mathbf{u}, \mathbf{l}}),$$

where the last step exploits homogeneity of the exponent measure. Transforming to pseudo-polar coordinates this becomes

$$v_X(\mathcal{C}_{\beta, \mathbf{u}, \mathbf{l}}) = \int_{\{(r, \mathbf{w}) : r\mathbf{w}_\beta > \mathbf{u}, r\mathbf{w}_{-\beta} < \mathbf{l}\}} \alpha r^{-\alpha-1} dr dH_X(\mathbf{w}).$$

Based on Eq. 10, there are only q possible angles along which extremes can occur, but only those with $\mathbf{a}_j / \|\mathbf{a}_j\|_\alpha \in C_\beta$ will contribute to the integral. To see this, consider $\mathbf{w} = \mathbf{a} / \|\mathbf{a}\|_\alpha \in C_\gamma$, where \mathbf{a} denotes an arbitrary column of A and $\gamma \in \mathcal{P}_d^*$. First, suppose $\gamma \subset \beta$ (strictly) so that there exists $i \in \beta$ such that $i \notin \gamma$. Then, for all $r > 0$, we have $r\mathbf{w}_\beta \not> \mathbf{u}$, since $rw_i = 0$. Similarly, suppose $\gamma \supset \beta$, in which case there exists $i \in \gamma$ with $i \notin \beta$. Let ℓ denote the entry of \mathbf{l} corresponding to component i . An extreme event along \mathbf{w} requires $rw_i < \ell$ and $r\mathbf{w}_\beta > \mathbf{u}$, but for \mathbf{u} sufficiently large these inequalities have no solution $r > 0$. On the other hand, if $\beta = \gamma$, then $\mathbf{w}_{-\beta} = \mathbf{0} \in \mathbb{R}^{d-s}$ and so

$$r\mathbf{w}_\beta > \mathbf{u}, r\mathbf{w}_{-\beta} < \mathbf{l} \iff r\mathbf{w}_\beta > \mathbf{u} \iff r > \max_{i=1, \dots, s} \left(\frac{\|\mathbf{a}\|_\alpha u_i}{a_{\beta_i, i}} \right).$$

This yields the final result

$$v_X(\mathcal{C}_{\beta, \mathbf{u}, \mathbf{l}}) = \sum_{j: \frac{\mathbf{a}_j}{\|\mathbf{a}_j\|_\alpha} \in C_\beta} \|\mathbf{a}_j\|_\alpha^\alpha \int_{\max_i \left(\frac{\|\mathbf{a}_j\|_\alpha u_i}{a_{\beta_i, j}} \right)}^\infty \alpha r^{-\alpha-1} dr = \sum_{j: \frac{\mathbf{a}_j}{\|\mathbf{a}_j\|_\alpha} \in C_\beta} \min_{i=1, \dots, s} \left(\frac{a_{\beta_i, j}}{u_i} \right)^\alpha.$$

For $\alpha = 1$ we can establish the upper bound $\hat{\mathbb{P}}(X \in \mathcal{C}_{\beta, \mathbf{u}, \mathbf{l}}) \leq H_X(C_\beta)/(su)$ with equality attained if and only if the angular measure places all of its mass on C_β at the centroid $\mathbf{e}(\beta)/|\beta|$, where $\mathbf{e}(\beta) := (\mathbb{1}\{i \in \beta\} : i = 1, \dots, d) \in \{0, 1\}^d$.

The form of this bound offers an intuitive interpretation as the limiting probability of the angular component lying in the required subspace multiplied by the survivor function of a Pareto($\alpha = 1$) random variable evaluated at the effective radial threshold $\|u\mathbf{1}_s\|_1 = su$.

3.1.3 Existing approaches to inference for max-linear models

Suppose $X \sim \text{MaxLinear}(A; \alpha)$, where α is known and A is to be estimated from a collection $\{\mathbf{x}_t = (r_t, \boldsymbol{\theta}_t) : t = 1, \dots, n\}$ of independent observations of $X = (R, \Theta)$. For $j = 1, \dots, n$, let $r_{(j)}$ denote the j th upper order statistic of $\{r_1, \dots, r_n\}$ and denote by $\mathbf{x}_{(j)}$ and $\boldsymbol{\theta}_{(j)}$ the corresponding observation vector and angular component, respectively. We consider two existing approaches for inferring the parameter matrix A : an empirical estimate and a CP-decomposition based estimate (Cooley and Thibaud 2019; Kiriliouk and Zhou 2022). The former is a natural estimate in view of Eq. 10; the latter exploits the connection between CP-factors and TPDMs outlined in the ensuing discussion.

Definition 4 For $1 \leq k < n$, the *empirical estimate* of A based on is $\hat{A} = (\hat{\mathbf{a}}_1, \dots, \hat{\mathbf{a}}_k) \in \mathbb{R}_+^{d \times k}$, where $\hat{\mathbf{a}}_j := (d/k)^{1/\alpha} \boldsymbol{\theta}_{(j)}$ for $j = 1, \dots, k$.

The quantity k is the customary tuning parameter representing the number of ‘extremes’ – observations with norm not less than the implied radial threshold $r_{(k)}$ – that enter into the estimator. The associated angular measure (for any α) and TPDM (for $\alpha = 2$) are given by

$$\hat{H}_X(\cdot) := H_{\hat{A} \circ \mathbf{Z}}(\cdot) = \frac{d}{k} \sum_{t=1}^n \mathbb{1}\{\boldsymbol{\theta}_t \in \cdot, r_t \geq r_{(k)}\}, \quad \hat{\Sigma} := \hat{A}\hat{A}^T.$$

These are the empirical angular measure (Einmahl and Segers 2009) and empirical TPDM (Cooley and Thibaud 2019), respectively. Due to non-uniqueness of CP-factors, further estimates of A can be obtained by CP-decomposition of the empirical TPDM.

Definition 5 Any CP-factor of $\hat{\Sigma}$ is called a *CP-estimate* of A , denoted \tilde{A} .

In general, \tilde{A} is distinct from \hat{A} and induces a different angular measure $\tilde{H}_X := H_{\tilde{A} \circ \mathbf{Z}}$. By construction, these angular measures give rise to identical tail pairwise dependence measures, since $\tilde{A}\tilde{A}^T = \hat{A}\hat{A}^T$. Note that \tilde{A} implicitly depends on the same tuning parameter k as \hat{A} via $\hat{\Sigma}$. In this paper, all CP-estimates will be obtained using the efficient factorisation procedure of Kiriliouk and Zhou (2022). Their algorithm takes as input a $d \times d$ TPDM with strictly positive entries and a permutation (i_1, \dots, i_d) of $\mathbb{V}(d)$ and returns a square CP-factor $\tilde{A} \in \mathbb{R}_+^{d \times d}$ whose columns satisfy $\tilde{\mathbf{a}}_j / \|\tilde{\mathbf{a}}_j\|_2 \in C_{\mathbb{V}(d) \setminus \{i_l : l < j\}}$ for $j = 1, \dots, d$. (Note that not all input permutations will result in a valid decomposition.)

3.1.4 Inference for max-linear models based on sparse projections

A limitation of \hat{A} and \tilde{A} is that they do not capture the extremal directions of X . For the empirical estimate, this is a consequence of $X/\|X\|_\alpha$ lying in the simplex interior $C_{\mathbb{V}(d)}$ with probability one. This implies that $\hat{\mathbb{P}}(\hat{A} \circ Z \in C_{\beta,u,l}) = 0$ for any $\beta \neq \mathbb{V}(d)$. On the other hand, the d extremal directions of $\tilde{A} \circ Z$ are fully determined by the user-defined input path (i_1, \dots, i_d) , namely $\mathbb{V}(d), \mathbb{V}(d) \setminus \{i_1\}, \mathbb{V}(d) \setminus \{i_1, i_2\}, \dots, \mathbb{V}(d) \setminus \{i_1, \dots, i_{d-1}\}$. Hence $\hat{\mathbb{P}}(\tilde{A} \circ Z \in C_{\beta,u,l}) > 0$ if and only if $\beta = \{i_{d-|\beta|+1}, \dots, i_d\}$. To address these shortcomings and better capture extremal directions, we propose augmenting the empirical estimate with an alternative notion of angle based on Euclidean projections onto the L_1 -simplex (Meyer and Wintenberger 2023).

Definition 6 The Euclidean projection onto the L_1 -simplex is defined by

$$\pi : \mathbb{R}_+^d \rightarrow \mathbb{S}_+^{d-1}, \quad \pi(v) = \operatorname{argmin}_{w \in \mathbb{S}_+^{d-1}} \|w - v\|_2^2.$$

This projection is useful because $\pi(v)$ may lie on a boundary of the simplex even if $v/\|v\|_1$ lies in its interior. Assume now that $\alpha = 1$.

Definition 7 For $1 \leq k < n$, the *sparse empirical estimate* of A is $\hat{A}^\star = (\hat{a}_1^\star, \dots, \hat{a}_k^\star) \in \mathbb{R}_+^{d \times k}$, where $\hat{a}_j^\star = dk^{-1}\pi(x_{(j)}/r_{(k+1)})$ for $j = 1, \dots, k$.

The corresponding model for the angular measure,

$$\hat{H}_X^\star(\cdot) := H_{\hat{A}^\star \circ Z}(\cdot) = \frac{d}{k} \sum_{j=1}^k \mathbb{1}\{\pi(x_{(j)}/r_{(k+1)}) \in \cdot\},$$

spreads mass across the subspaces $C_\beta \subseteq \mathbb{S}_+^{d-1}$ in which the projected data lie and hence $\hat{\mathbb{P}}(\hat{A}^\star \circ Z \in C_{\beta,u}) \neq 0$ for all corresponding β . A full study of the theoretical properties of \hat{A}^\star has not been conducted. Roughly speaking, according to the theory in Meyer and Wintenberger (2023), the angular measure associated with \hat{A}^\star will place mass on the correct subspaces, but may additionally incorrectly assign mass to lower-dimensional ('non-maximal') faces. Moreover, the estimator will perform best in the idealised case where the classical and sparse notions of regular variation coincide – see Corollary 1 in Meyer and Wintenberger (2023). Beyond this, presently we cannot provide rigorous statistical guarantees for our approach. Having introduced our estimator and all the requisite theory, we are ready to present our methods for the multivariate challenges.

3.2 Challenge 3

3.2.1 Data

Challenge 3 considers a trivariate random vector $Y = (Y_1, Y_2, Y_3)$ on standard Gumbel margins, i.e. $\mathbb{P}(Y_i < y) = G(y) := \exp(-\exp(-y))$ for $y \in \mathbb{R}$ and $i = 1, 2, 3$. It

entails estimating

$$p_1 := \mathbb{P}(Y_1 > 6, Y_2 > 6, Y_3 > 6), \quad p_2 := \mathbb{P}(Y_1 > 7, Y_2 > 7, Y_3 < -\log(\log(2))).$$

The data comprise $n = 21,000$ independent observations $\{y_t = (y_{t1}, y_{t2}, y_{t3}) : t = 1, \dots, n\}$ of \mathbf{Y} . The available covariate information is not leveraged by our method.

3.2.2 Methodology

Let $\mathbf{X} = (X_1, X_2, X_3)$ denote the random vector obtained by transforming \mathbf{Y} to Fréchet margins with shape parameter 1, i.e. $X_i = \Psi_1^{-1}(G(Y_i)) = \exp(Y_i) \sim \Psi_1$ for $i = 1, 2, 3$. The above probabilities can be expressed as

$$p_1 = \mathbb{P}(X_1 > e^6, X_2 > e^6, X_3 > e^6), \quad p_2 = \mathbb{P}\left(X_1 > e^7, X_2 > e^7, X_3 < \frac{1}{\log 2}\right). \quad (12)$$

The thresholds e^6 , e^7 , and $1/\log(2)$ correspond approximately to the 99.8%, 99.9% and 50% quantiles of Ψ_1 , respectively. Our solution models \mathbf{X} as max-linear, infers A using the sparse empirical estimator (for some hyperparameter k) and computes estimates for p_1 and p_2 via (11), that is

$$\hat{p}_1 = \hat{\mathbb{P}}(\hat{A}^* \circ \mathbf{Z} \in \mathcal{C}_{\mathbb{V}(3), e^6}), \quad \hat{p}_2 = \hat{\mathbb{P}}(\hat{A}^* \circ \mathbf{Z} \in \mathcal{C}_{\{1,2\}, e^7, 1/\log(2)}). \quad (13)$$

Inference is based on the transformed data $\{\mathbf{x}_t = (x_{t1}, x_{t2}, x_{t3}) : t = 1, \dots, n\}$, where $x_{ti} := \exp(y_{ti})$ for $t = 1, \dots, n$ and $i = 1, 2, 3$.

3.2.3 Results

The results presented are based on $k = 500 \approx 2.5\% \times n$. This value was used for our competition submission and the results' sensitivity to this choice will be examined later.

The ternary plots in Fig. 6 depict the angular components associated with the k largest observations in norm, i.e. exceeding the radial threshold $r_{(k+1)} \approx 138.77$. The left-hand plot represents the self-normalised vectors $\theta_{(1)}, \dots, \theta_{(k)}$. We find points lying near the centre of the triangle and in the neighbourhood of all its edges and vertices. This suggests that the angular measure spreads mass across all seven faces of \mathbb{S}_+^2 . However, we reiterate that no points lie *exactly* on the boundary. In contrast, the sparse angles $\{\pi(\mathbf{x}_t/r_{(k+1)}) : r_t > r_{(k+1)}\}$ in the right-hand plot lie in the interior (black, 40 points), along the edges (red, 139 points), and on the vertices (blue, 321 points) of the closed simplex. Only the 40 vectors in $C_{\mathbb{V}(3)}$ and 23 vectors in $C_{\{1,2\}}$ will enter into the estimates of Eq. 12.

The projected vectors are collated to form the 3×500 matrix \hat{A}^* . The first 100 columns of the matrices \hat{A} and \hat{A}^* are represented visually in Fig. 7. As expected, \hat{A} is

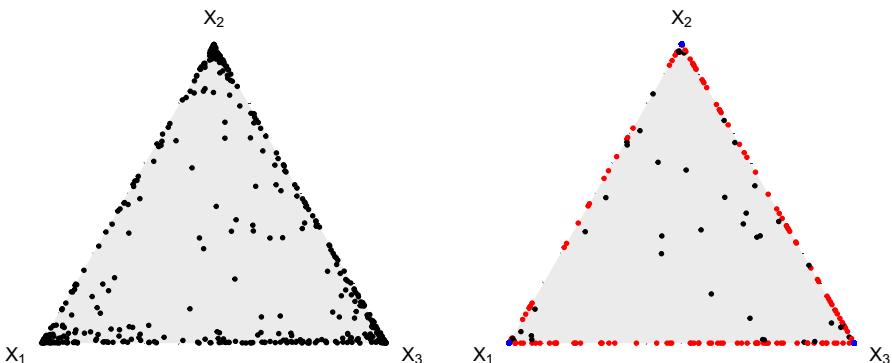


Fig. 6 The angles $x_{(j)}/\|x_{(j)}\|_1 \in S^2_+$ (left) and Euclidean projections $\pi(x_{(j)}/r_{(k+1)}) \in S^2_\perp$ (right) for the 500 largest observations in Challenge 3. Points are coloured according to whether they lie in the interior (black), on an edge (red) or on a vertex (blue)

dense, albeit populated with small values, while \hat{A}^* exhibits a high degree of sparsity, in the sense that most of its columns satisfy $\|\hat{a}_j^*\|_0 < \|\hat{a}_j\|_0 = 3$. Duplicated columns in \hat{A}^* could be merged and re-weighted to produce a compressed estimate \hat{A}_{comp}^* with $q_{\text{comp}} = 40 + 139 + 3 = 182$ unique columns, but ultimately they parametrise identical models since $H_{\hat{A}^* \circ Z} = H_{\hat{A}_{\text{comp}}^* \circ Z}$.

Substituting \hat{A}^* into Eq. 13 yields the final point estimates $\hat{p}_1 = 3.36 \times 10^{-5}$ and $\hat{p}_2 = 2.76 \times 10^{-5}$, to three significant figures. We are pleased to find that our estimates are very close to the true values given in Rohrbeck et al. (2023). Furthermore, Fig. 8 shows that our estimates of p_2 are fairly stable across a range of ‘reasonable’ choices of k , say, $1\% \leq k/n \leq 5\%$. When k is smaller than this, the estimates become highly variable due to the limited effective sample size. If k is too large, we risk introducing bias by including observations that do not reflect the true tail dependence structure. The estimates of p_1 exhibit greater sensitivity to the choice of threshold. The development of theoretically justified diagnostics/procedures for selecting the threshold represents an avenue for future work.

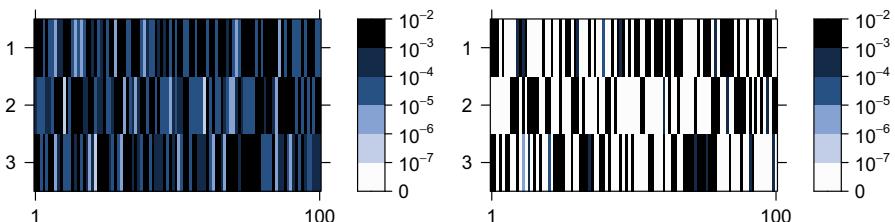


Fig. 7 The first 100 columns of \hat{A} (left) and \hat{A}^* (right) for Challenge 3. The colour intensity of each cell represents the magnitude of the corresponding matrix entry

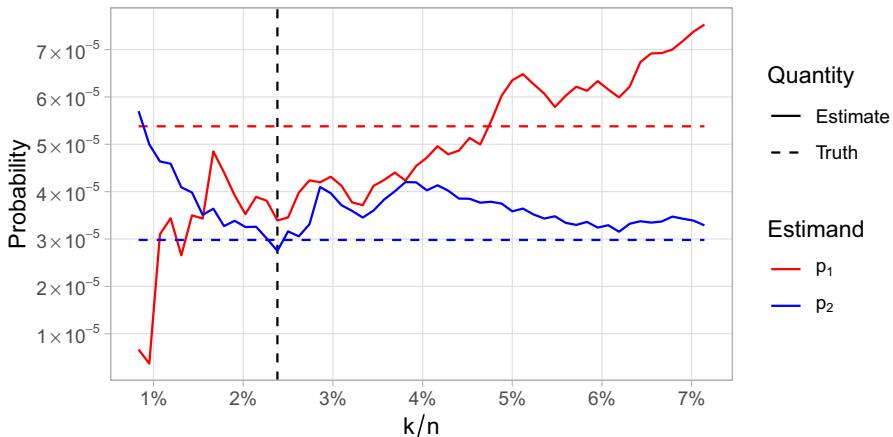


Fig. 8 Challenge 3 results for different choices for the tuning parameter, k . The horizontal dashed lines indicate the estimands' true values. Our submitted solution was based on $k = 500$ (vertical dashed line)

3.3 Challenge 4

3.3.1 Data

Challenge 4 regards a $d = 50$ dimensional random vector \mathbf{Y} on standard Gumbel margins. The components of \mathbf{Y} are random variables $Y_{i,j}$ for $i = 1, \dots, 25$ and $j = 1, 2$, representing the value of an environmental variable at the i th site in the administrative area of government area j . The joint exceedance probabilities

$$\begin{aligned} p_1 &:= \mathbb{P}(Y_{i,j} > G^{-1}(1 - \phi_j) : i = 1, \dots, 25, j = 1, 2), \\ p_2 &:= \mathbb{P}(Y_{i,j} > G^{-1}(1 - \phi_1) : i = 1, \dots, 25, j = 1, 2), \end{aligned}$$

where $\phi_1 = 1/300$ and $\phi_2 = 12/300$, are to be estimated from $n = 10,000$ independent observations $\{\mathbf{y}_t = (y_{t,i,j} : i = 1, \dots, 25, j = 1, 2) : t = 1, \dots, n\}$ of \mathbf{Y} .

3.3.2 Methodology

Let $\mathbf{X} = (X_1, \dots, X_d)$ denote the random vector obtained from \mathbf{Y} by re-indexing its variables and transforming to Fréchet margins with shape parameter $\alpha = 2$, i.e.

$$X_{i+25(j-1)} := \Psi_2^{-1}(G(Y_{i,j})) = \exp(Y_{i,j}/2) \sim \Psi_2, \quad (i = 1, \dots, 25, j = 1, 2).$$

The choice of $\alpha = 2$ will be justified later. The data are transformed in an identical way yielding $\{\mathbf{x}_t = (x_{t,1}, \dots, x_{t,d}) : t = 1, \dots, n\}$, where $x_{t,i+25(j-1)} := \exp(y_{t,i,j}/2)$ for

$t = 1, \dots, n, i = 1, \dots, 25$ and $j = 1, 2$. The estimands can now be expressed as

$$p_1 = \mathbb{P}(X \in \mathcal{C}_{\mathbb{V}(d), \mathbf{u}_1}), \quad [\mathbf{u}_1]_i = \begin{cases} \Psi_2^{-1}(1 - \phi_1), & \text{if } 1 \leq i \leq 25 \\ \Psi_2^{-1}(1 - \phi_2), & \text{if } 26 \leq i \leq 50 \end{cases},$$

$$p_2 = \mathbb{P}(X \in \mathcal{C}_{\mathbb{V}(d), \mathbf{u}_2}), \quad \mathbf{u}_2 = \Psi_2^{-1}(1 - \phi_1)\mathbf{1}_{50}.$$

At a high level, our method proceeds in a similar vein to Challenge 3: we will model X as max-linear and compute estimates for p_1 and p_2 based on Eq. 11. However, our exploratory analysis revealed that not all components of X are asymptotically dependent, implying that $H_X(\mathbb{V}(d)) = 0$. This has important ramifications for how we construct our solution. On the one hand, empirical or CP-estimates of A will assert that $C_{\mathbb{V}(d)}$ is an extremal direction, resulting in a misspecified model. On the other hand, a model that correctly identifies $C_{\mathbb{V}(d)}$ as a H_X -null set will yield $\hat{p}_1 = \hat{p}_2 = 0$, despite the true values being non-zero (Rohrbeck et al. 2023). How do we resolve this difficulty? The key is to identify clusters of asymptotically dependent variables prior to model fitting and estimate the probability of concurrent joint exceedance events in all clusters. Our working hypothesis – that the marginal variables can be partitioned such that asymptotic independence is present between clusters but not within them – is formalised below.

Assumption 1 There exists $2 \leq K \leq d$ and a partition β_1, \dots, β_K of $\mathbb{V}(d)$ such that the angular measure is supported on the closed subspaces $\bar{C}_{\beta_1}, \dots, \bar{C}_{\beta_K} \subset \mathbb{S}_+^{d-1}$, where $\bar{C}_\beta := \{C_{\beta'} : \beta' \subseteq \beta\}$ for any $\beta \in \mathcal{P}_d^*$. That is, $H_X(\cup_{l=1}^K \bar{C}_{\beta_l}) = H_X(\mathbb{S}_+^{d-1})$.

This scenario has been considered before, cf. Assumption 1 in Fomichov and Ivanovs (2023). If X is max-linear with parameter $A = (\mathbf{a}_1, \dots, \mathbf{a}_q) \in \mathbb{R}_+^{d \times q}$, we may restate the assumption as follows.

Assumption 2 There exist permutations $\pi : \mathbb{V}(d) \rightarrow \mathbb{V}(d)$ and $\phi : \mathbb{V}(q) \rightarrow \mathbb{V}(q)$ such that $X_\pi := (X_{\pi(1)}, \dots, X_{\pi(d)}) \sim \text{MaxLinear}(A_\phi; \alpha)$, where $A_\phi := (\mathbf{a}_{\phi(1)}, \dots, \mathbf{a}_{\phi(q)}) \in \mathbb{R}_+^{d \times q}$ is block-diagonal with $2 \leq K \leq d$ blocks. For $l = 1, \dots, K$, the l th block matrix $A_\phi^{(l)}$ has $d_l = |\beta_l|$ rows, $1 \leq q_l < q$ columns, and is such that the $q_l \times q_l$ matrix $A_\phi^{(l)}(A_\phi^{(l)})^T$ has strictly positive entries. The blocks' dimensions satisfy $\sum_{l=1}^K d_l = d$ and $\sum_{l=1}^K q_l = q$.

Under this framework, X divides into random sub-vectors $X^{(1)}, \dots, X^{(K)}$, where $X^{(l)} := (X_j : j \in \beta_l) \sim \text{MaxLinear}(A_\phi^{(l)}; \alpha)$ for $l = 1, \dots, K$. (Henceforth, assume for notational convenience that the columns of A are already appropriately ordered, so that $A = A_\phi$.) While the clustering assumption is simplistic and cannot be expected to hold in general applications, it reflects what we suspect to be the true dependence structure for Challenge 4. Moreover, it is dimension reducing because the original d -dimensional problem is transformed to a set of K independent problems with dimensions $d_1, \dots, d_K < d$. This ameliorates the curse of dimensionality to some extent.

In general, a joint exceedance event can be decomposed into concurrent joint exceedances in all K clusters as $\{X \in \mathcal{C}_{\mathbb{V}(d), \mathbf{u}}\} = \cap_{l=1}^K \{X \in \mathcal{C}_{\mathbb{V}(d_l), \mathbf{u}^{(l)}}\}$, where each

threshold sub-vector $\mathbf{u}^{(l)}$ is defined from \mathbf{u} analogously to $\mathbf{X}^{(l)}$. Since we consider variables in different clusters to be asymptotically independent, we assume that for large, finite thresholds, joint exceedances in each cluster are approximately independent events, so that

$$\mathbb{P}(\mathbf{X} \in \mathcal{C}_{\mathbb{V}(d), \mathbf{u}}) = \mathbb{P}\left(\bigcap_{l=1}^K \{\mathbf{X}^{(l)} \in \mathcal{C}_{\mathbb{V}(d_l), \mathbf{u}^{(l)}}\}\right) \approx \prod_{l=1}^K \mathbb{P}(\mathbf{X}^{(l)} \in \mathcal{C}_{\mathbb{V}(d_l), \mathbf{u}^{(l)}}).$$

Assuming $\mathbf{X}^{(l)} \sim \text{MaxLinear}(A^{(l)}; \alpha)$ for $l = 1, \dots, K$, each term in the product can be estimated using Eq. 11, so that

$$\mathbb{P}(\mathbf{X} \in \mathcal{C}_{\mathbb{V}(d), \mathbf{u}}) \approx \prod_{l=1}^K \hat{\mathbb{P}}(A^{(l)} \circ \mathbf{Z} \in \mathcal{C}_{\mathbb{V}(d_l), \mathbf{u}^{(l)}}). \quad (14)$$

The final step is to replace $A^{(1)}, \dots, A^{(K)}$ with suitable estimates. We opted to use CP-estimates for two reasons: (i) they are rooted in the TPDM, which is geared towards high-dimensional settings, and (ii) their non-uniqueness enables us to compute numerous parameter/probability estimates, whose variation reflects the model uncertainty that arises from summarising dependence via the TPDM and thereby overlooking higher-order dependencies between components. The use of CP-estimates justifies our choosing $\alpha = 2$ in the pre-processing step and throughout.

3.3.3 Results

We now present our results for Challenge 4. First, the variables X_1, \dots, X_d are partitioned into K groups based on asymptotic (in)dependence using the clustering algorithm of Bernard et al. (2013). This entails constructing a distance matrix $\mathcal{D} = (\hat{d}_{ij})$, where \hat{d}_{ij} denotes a non-parametric estimate of the F-madogram distance between variables X_i and X_j . The distance metric is connected to the strength of extremal dependence between X_i and X_j , with $\hat{d}_{ij} \approx 0$ implying strong asymptotic dependence and $\hat{d}_{ij} = 1/6$ in the case of asymptotic independence. The partition around medoids (PAM) clustering algorithm (Kaufman and Rousseeuw 1990) returns a partition β_1, \dots, β_K of $\mathbb{V}(d)$ based on \mathcal{D} . The number of clusters K is a pre-specified tuning parameter; we identify $K = 5$ clusters whose sizes are given in Table 1. By consulting Rohrbeck et al. (2023), we have verified that our clustering was correct. Defining cluster membership variables $\mathcal{M}_1, \dots, \mathcal{M}_d \in \{1, \dots, K\}$ by $\mathcal{M}_i = l \iff i \in \beta_l$ for $i = 1, \dots, d$, we find that $\max\{\hat{d}_{ij} : \mathcal{M}_i = \mathcal{M}_j\} = 0.113 < 1/6$ and $\min\{\hat{d}_{ij} : \mathcal{M}_i \neq \mathcal{M}_j\} = 0.164 \approx 1/6$. These summary statistics are consistent with Assumption 1.

Next, we compute the empirical TPDMs of $\mathbf{X}^{(1)}, \dots, \mathbf{X}^{(K)}$. For $l = 1, \dots, K$ and $t = 1, \dots, n$, define the observational sub-vector $\mathbf{x}_t^{(l)} = (x_{ti} : i \in \beta_l)$ and its radial and angular components $r_t^{(l)} = \|\mathbf{x}_t^{(l)}\|_2, \theta_t^{(l)} = \mathbf{x}_t^{(l)} / \|\mathbf{x}_t^{(l)}\|_2$, respectively. Let $\mathbf{x}_{(j)}^{(l)}, r_{(j)}^{(l)}$ and $\theta_{(j)}^{(l)}$ denote the vector, radius, and angle associated with the j th largest observation

Table 1 Summary statistics for the Challenge 4 clusters and their empirical TPDMs

Cluster, l	Size, d_l	U1 sites	U2 sites	$\left\{ \hat{\sigma}_{X_{ij}^{(l)}} : i \neq j \right\}$	Min.	Median	Max.
1	9	7	2	0.30	0.33	0.40	
2	8	5	3	0.64	0.68	0.74	
3	8	3	5	0.62	0.67	0.74	
4	12	7	5	0.27	0.33	0.39	
5	13	3	10	0.43	0.50	0.58	

in norm among $\mathbf{x}_1^{(l)}, \dots, \mathbf{x}_n^{(l)}$. Choose a tuning parameter $1 \leq k_l \leq n$ representing the number of extreme observations that enter into the estimate for cluster l . Then

$$\hat{A}^{(l)} = \left(\frac{d_l}{k_l} \boldsymbol{\theta}_{(1)}^{(l)}, \dots, \frac{d_l}{k_l} \boldsymbol{\theta}_{(k_l)}^{(l)} \right), \quad \hat{\Sigma}_{X^{(l)}} = \hat{A}^{(l)} (\hat{A}^{(l)})^T.$$

We set $k_1 = \dots = k_K =: k = 250$, corresponding to a sampling fraction of $k/n = 2.5\%$ for each cluster. The empirical TPDMs for the first two clusters are displayed in Fig. 9; summary statistics for all clusters' TPDMs are listed in Table 1. Asymptotic dependence is strongest in clusters 2 and 3 and weakest in clusters 1 and 4.

By repeated application of the CP-decomposition algorithm of Kiriliouk and Zhou (2022) with randomly chosen inputs, we obtain $N_{\text{cp}} = 50$ CP-estimates of each matrix $A^{(l)}$. The resulting CP-factors are denoted by $\tilde{A}_1^{(l)}, \dots, \tilde{A}_{N_{\text{cp}}}^{(l)}$. Note that among $\tilde{A}_1^{(l)}, \dots, \tilde{A}_{N_{\text{cp}}}^{(l)}$ there are at most d_l distinct leading columns, because there are only d_l unique ways to initialise the (deterministic) algorithm. But $\hat{\mathbb{P}}(\tilde{A} \circ \mathbf{Z} \in \mathcal{C}_{\mathbb{V}(d), \mathbf{u}})$ is fully determined by $\tilde{\mathbf{a}}_1$, so in fact $\{\hat{\mathbb{P}}(\tilde{A}_r^{(l)} \circ \mathbf{Z} \in \mathcal{C}_{\mathbb{V}(d_l), \mathbf{u}^{(l)}}) : r = 1, \dots, N_{\text{cp}}\}$ contains at most d_l distinct values. These values are represented by black points in the left-hand plot in Fig. 10. Clusters 2 and 3 are deemed most likely to experience a joint extreme

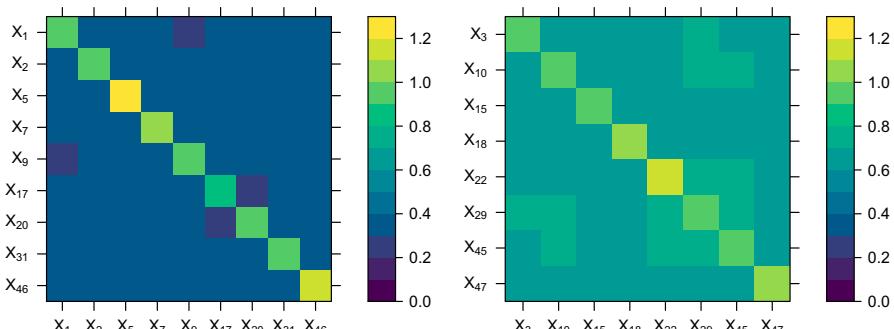


Fig. 9 The estimated TPDMs for the first two clusters, $\hat{\Sigma}_{X^{(1)}}$ (left) and $\hat{\Sigma}_{X^{(2)}}$ (right), based on the $k = 250$ most extreme observations in norm

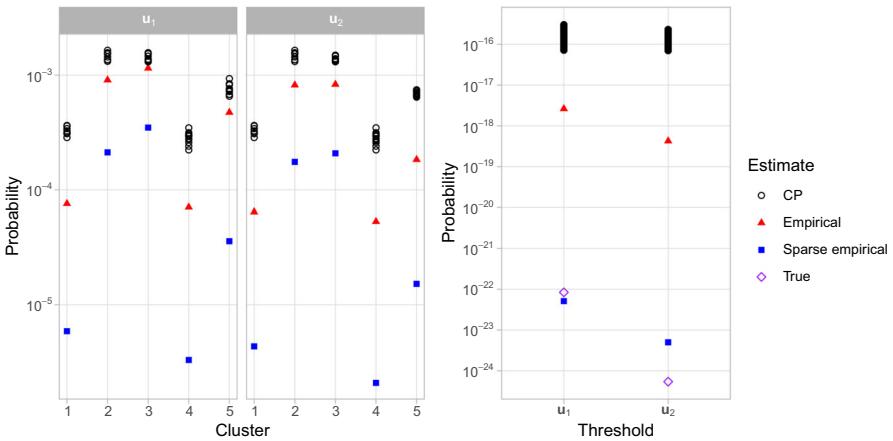


Fig. 10 Joint exceedance probability estimates for each cluster sub-vector (left) and the full vector (right) based on different estimators for the parameters $A^{(1)}, \dots, A^{(5)}$

event, because they contain a small number ($d_2 = d_3 = 8$) of highly dependent variables. The effect of changing the threshold from \mathbf{u}_1 to \mathbf{u}_2 is most pronounced in cluster 5, since it is primarily composed of sites in area U2.

Substituting each CP-estimate into Eq. 14 and enumerating over all possible combinations, we produce sets of estimates of p_i , for $i = 1, 2$, given by

$$\tilde{P}_i := \{\hat{\mathbb{P}}(\tilde{A}_{r_1}^{(1)} \circ \mathbf{Z} \in \mathcal{C}_{\mathbb{V}(d_1), \mathbf{u}_i^{(1)}}) \times \dots \times \hat{\mathbb{P}}(\tilde{A}_{r_K}^{(K)} \circ \mathbf{Z} \in \mathcal{C}_{\mathbb{V}(d_K), \mathbf{u}_i^{(K)}}) : r_1, \dots, r_K = 1, \dots, N_{\text{cp}}\}.$$

Each set has size $N_{\text{cp}}^K \approx 3 \times 10^8$ (including repeated values) and contains $\prod_{l=1}^K d_l = 89,856$ distinct values. The distributions of the estimates are represented by the black points in the right-hand panel of Fig. 10. Our final point estimates are taken as the median values $\tilde{p}_1 := \text{median}(\tilde{P}_1) = 1.4 \times 10^{-16}$ and $\tilde{p}_2 := \text{median}(\tilde{P}_2) = 1.3 \times 10^{-16}$, respectively, to two significant figures.

3.3.4 Improving performance using sparse empirical estimates

Unfortunately, it transpires that our method dramatically over-estimated the true probabilities. In hindsight, this could have been anticipated in view of the simulation study in Section 5.1 in Kiriliouk and Zhou (2022), where the authors remark that failure regions of the type $\mathcal{C}_{\mathbb{V}(d), \mathbf{u}}$ are poorly summarised by the TPDM. This prompts us to investigate whether using empirical or sparse empirical estimates instead of a CP-based approach would have improved our performance. For the former, this simply involves replacing $A^{(l)}$ with the precomputed matrix $\hat{A}^{(l)}$ in 14. The approach based on sparse empirical estimates proceeds analogously except that we must revert to the $\alpha = 1$ setting and transform the data and thresholds accordingly. The resulting estimates for the joint exceedance probabilities are represented by the red and blue points in Fig. 10. The values are smaller (better) in all clusters and consequently the final

estimates are closer to the true event probabilities (purple). In fact, using sparse empirical estimates yields $\hat{p}_1^* = 5.1 \times 10^{-23}$ and $\hat{p}_2^* = 5.0 \times 10^{-24}$, which are remarkably close to the correct solutions $p_1 = 8.4 \times 10^{-23}$ and $p_2 = 5.4 \times 10^{-25}$. Submitting these values would have significantly improved our ranking for this sub-challenge.

4 Conclusion

We propose four different methods in order to solve the EVA (2023) Conference Data Challenge. These methods allow for the estimation of extreme quantiles and probabilities. In the univariate cases, we estimate the conditional quantile by attempting to cluster the covariates to categorise the extremes. Once clustered, bootstrapping was used in order to estimate central 50% confidence intervals of the extreme quantiles. In practice however, we found no clearly defined clusters and that the Gaussian mixture distribution was not able to capture the non-Gaussian distribution of the covariates. This could also be extended by using GPDs with covariate models for each cluster, to try to improve quantile estimates. In the second challenge, we were able to assess the GPD fit of a misspecified model and apply a correction to an initial fit where there was large uncertainty in the tail behaviour. This resulted in a change in the return level from an under-estimate to a potentially less-severe over-estimate. By using a more sophisticated model for the covariates, further work can be done to articulate the extent to which the correction was due to a bias in the model fit.

Our performance in the multivariate challenges demonstrates that the max-linear model provides a good framework for estimating tail event probabilities. By connecting this model with sparse simplex projections, one can achieve exceptional performance on both Challenges 3 and 4. Given these results, further research on the theoretical properties of the sparse empirical estimator \hat{A}^* is warranted. An obvious shortcoming of our Challenge 3 methodology is that it ignores the available covariate information. Incorporating covariates into the max-linear/TPDM framework remains an unexplored area. The upwards bias in our Challenge 4 probability estimates is partly caused by over-estimating the dependence strength between certain groups of variables. The empirical TPDM has a known tendency to over-estimate weak/moderate dependence (Fix et al. 2021), as is the case in clusters 1, 3, and 5. A better TPDM estimator that takes this issue into account would likely improve our final results. In Challenges 3 and 4 we switch between $\alpha = 1$ and $\alpha = 2$ in a way that is somewhat unsatisfactory (mathematically and presentationally). In hindsight, employing a generalised version of the TPDM (Kiriliouk and Zhou 2022, Equation 4) would have allowed us to fix $\alpha = 1$ throughout. Finally, uncertainty quantification is a crucial aspect of risk assessment in practice, especially when dealing with the exceedingly rare events encountered in Challenge 4. Fix et al. (2021) utilise a bootstrapping procedure in the context of TPDM-based inference for extremes and their approach would transfer to our methods very naturally.

Acknowledgements The authors are thankful to Christian Rohrbeck, Emma Simpson and Jonathan Tawn of the University of Bath, University College London, and Lancaster University, respectively, for organising the challenges that inspired this work.

Author Contributions Both authors contributed equally to this manuscript. H.E. researched and wrote the content in Section 2 (univariate challenges). M.P. researched and wrote the content in Section 3 (multivariate challenges). Both authors contributed equally to the introduction and conclusion sections. Both authors reviewed the manuscript.

Funding Matthew Pawley is supported by a scholarship from the EPSRC Centre for Doctoral Training in Statistical Applied Mathematics at Bath (SAMBa), under the project EP/S022945/1. Henry Elsom is supported by the University Research Studentship Award (URSA) from the University of Bath.

Data Availability The data for the EVA (2023) Conference Data Challenge has been made publicly available by Rohrbeck et al. (2024+). The code used for Challenges 3 and 4 is available at <https://github.com/pawleymatthew/eva-2023-data-challenge>.

Declarations

Ethical Approval Not Applicable.

Competing Interests The authors declare no competing interests.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Bernard, E., Naveau, P., Vrac, M., Mestre, O.: Clustering of maxima: Spatial dependencies among heavy rainfall in France. *J. Clim.* **26**(20), 7929–7937 (2013)
- Clarkson, D., Eastoe, E., Leeson, A.: The importance of context in extreme value analysis with application to extreme temperatures in the U.S. and Greenland. *J. Royal Stat. Soc. Ser. C Appl. Stat.* **72**(4), 829–843 (2023)
- Coles, S.: An Introduction to Statistical Modeling of Extreme Values, 1st edn. Springer, London (2001)
- Coles, S.G., Powell, E.A.: Bayesian methods in extreme value modelling: A review and new developments. *Int. Stat. Rev.* **64**(1), 119–136 (1996)
- Cooley, D., Thibaud, E.: Decompositions of dependence for high-dimensional extremes. *Biometrika* **106**(3), 587–604 (2019)
- Davison, A.C., Smith, R.L.: Models for exceedences over high thresholds. *J. Roy. Stat. Soc.: Ser. B (Methodol.)* **52**(3), 393–425 (1990)
- Efron, B., Stein, C.: The Jackknife Estimate of Variance. *Ann. Stat.* **9**(3) (1981)
- Einmahl, J.H.J., Segers, J.: Maximum empirical likelihood estimation of the spectral measure of an extreme-value distribution. *Ann. Stat.* **37**(5B) (2009)
- Engelke, S., Ivanovs, J.: Sparse Structures for Multivariate Extremes. *Annu. Rev. Stat. Appl.* **8**(1), 241–270 (2021)
- Farkas, S., Lopez, O., Thomas, M.: Cyber claim analysis using generalized Pareto regression trees with applications to insurance. *Insur. Math. Econ.* **98**, 92–105 (2021)
- Fawcett, L., Walshaw, D.: A hierarchical model for extreme wind speeds. *J. Roy. Stat. Soc.: Ser. C (Appl. Stat.)* **55**(5), 631–646 (2006)
- Fix, M.J., Cooley, D., Thibaud, E.: Simultaneous autoregressive models for spatial extremes. *Environmetrics* **32**(2) (2021)

- Fomichov, V., Ivanovs, J.: Spherical clustering in detection of groups of concomitant extremes. *Biometrika* **110**(1), 135–153 (2023)
- Fougères, A.-L., Mercadier, C., Nolan, J.P.: Dense classes of multivariate extreme value distributions. *J. Multivar. Anal.* **116**, 109–129 (2013)
- Fraley, C., Raftery, A.: Model-based clustering, discriminant analysis, and density estimation. *J. Am. Stat. Assoc.* **97**(458), 611–631 (2002)
- Fraley, C., Raftery, A.: Enhanced model-based clustering, density estimation, and discriminant analysis software: MCLUST. *J. Classif.* **20**, 263–286 (2003)
- Gelfand, A., Smith, A.: Sampling-based approaches to calculating marginal densities. *J. Am. Stat. Assoc.* **85**(410), 398–409 (1990)
- Gilleland, E.: Bootstrap Methods for Statistical Inference. Part II: Extreme-Value Analysis. *J. Atmos. Oceanic Technol.* **37**, 2135–2144 (2020)
- Goix, N., Sabourin, A., Cléménçon, S.: Sparse representation of multivariate extremes with applications to anomaly detection. *J. Multivar. Anal.* **161**, 12–31 (2017)
- Gouldsborough, L., Hossaini, R., Eastoe, E., Young, P.J.: A temperature dependent extreme value analysis of UK surface ozone, 1980–2019. *Atmospheric Environment* **273** (2022)
- Hastie, T., Tibshirani, R., Friedman, J.: *The Elements of Statistical Learning: Data Mining, Inference and Prediction*. Springer, New York (2009)
- Katz, R.W.: Extreme value theory for precipitation: sensitivity analysis for climate change. *Adv. Water Resour.* **23**(2), 133–139 (1999)
- Katz, R.W., Parlange, M.B., Naveau, P.: Statistics of extremes in hydrology. *Adv. Water Resour.* **25**(8), 1287–1304 (2002)
- Kaufman, L., Rousseeuw, P.J.: *Finding Groups in Data*. Wiley Series in Probability and Statistics. John Wiley & Sons Inc, Hoboken, NJ, USA (1990)
- Kiriliouk, A.: Hypothesis testing for tail dependence parameters on the boundary of the parameter space. *Econometrics and Statistics* **16**, 121–135 (2020)
- Kiriliouk, A., Zhou, C.: Estimating probabilities of multivariate failure sets based on pairwise tail dependence coefficients (2022). Preprint at <https://arxiv.org/abs/quant-ph/0208066v1>
- Klüppelberg, C., Krali, M.: Estimating an extreme Bayesian network via scalings. *J. Multivar. Anal.* **181**, 104672 (2021)
- Kunz, M., Mohr, S., Rauthé, M., Lux, R., Kottmeier, C.: Assessment of extreme wind speeds from Regional Climate Models - Part 1: Estimation of return values and their evaluation. *Nat. Hazards Earth Syst. Sci.* **10**, 907–922 (2010)
- Larsson, M., Resnick, S.: Extremal dependence measure and extremogram: the regularly varying case. *Extremes* **15**(2), 231–256 (2012)
- McLachlan, G., Peel, D.: *Finite Mixture Models*. Wiley Series in Probability and Statistics, John Wiley & Sons, Inc., New York (2000)
- Meyer, N., Wintenberger, O.: Multivariate sparse clustering for extremes. *J. Am. Stat. Assoc.* 1–12 (2023)
- Mouret, F., Albughdadi, M.: Duthoit, S.: Kouamé, D.: Tourneret, J.-Y.: Robust Estimation of Gaussian Mixture Models Using Anomaly Scores and Bayesian Information Criterion for Missing Value Imputation, 30th European Signal Processing Conference (EUSIPCO), 827–831 (2022)
- Pickands, J.: Statistical inference using extreme order statistics. *Ann. Stat.* **3**(1), 119–131 (1975)
- Resnick, S.: *Heavy-tail Phenomena: Probabilistic and Statistical Modeling*. Springer series in operations research and financial engineering. Springer, New York, N.Y (2007)
- Rohrbeck, C., Simpson, E.S., Tawn, J.A.: Editorial: EVA Conference Data Challenge (2023)
- Rohrbeck, C., Simpson, E.S., Tawn, J.A.: Dataset for EVA 2023 Data Challenge. Bath: University of Bath Research Data Archive (2024). <https://doi.org/10.15125/BATH-01399>
- Simpson, E.S., Wadsworth, J.L., Tawn, J.A.: Determining the dependence structure of multivariate extremes. *Biometrika* **107**(3), 513–532 (2020)
- Towler, E.L., Rajagopalan, B., Gilleland, E., Summers, R.S., Yates, D., Katz, R.W.: Modeling hydrologic and water quality extremes in a changing climate: A statistical approach based on extreme value theory. *Water Resources Research* **46**(11), (2010)