Extremes
https://doi.org/10.1007/s10687-026-00549-4

RESEARCH

Univariate modelling of multivariate precipitation extremes using the block quantiles method

EVA (2025) data challenge: Team Moscow State

Ruslan Guryev[1] · Valentin Kuzmenko[1] · Stepan Platinskii[2] · Idar Saralpov[1] · Igor Rodionov[3]

Received: 6 January 2026 / Revised: 24 August 2026 / Accepted: 1 September 2026
© The Author(s) 2026

Abstract
This paper presents the approach of Team Moscow State to solving the tasks of the EVA (2025) Conference Data Challenge. Although the challenge tasks were originally formulated in a multivariate setting, they can be reformulated as problems of univariate statistics of extremes, which largely determined our approach to their solution. It relies on a unified framework that combines classical techniques of statistics of extremes with the recently proposed block quantiles method.

Keywords Block quantiles method · Generalised Pareto distribution · Precipitation extremes · Statistics of extremes

Mathematics Subject Classification (2010) 62G32 · 62P12

✉ Igor Rodionov
igor.rodionov@essex.ac.uk

Ruslan Guryev
ruslan.gurev@math.msu.ru

Valentin Kuzmenko
valentin.kuzmenko@math.msu.ru

Stepan Platinskii
platinsky.sa@gmail.com

Idar Saralpov
idar.saralpov@math.msu.ru

1 Faculty of Mechanics and Mathematics, Lomonosov Moscow State University, Leninskie Gory, Moscow 119991, Russia

2 Yandex LLC, Ulitsa Lva Tolstogo 16, Moscow 119021, Russia

3 School of Mathematics, Statistics and Actuarial Science, University of Essex, Wivenhoe Park, Colchester CO43SQ, UK

Published online: 10 September 2026

1 Introduction

Estimating tail characteristics is essential in domains where extreme events with potentially catastrophic consequences may occur. Many of such catastrophes are natural phenomena, such as strong winds that damage infrastructure, severe droughts or heavy rainfall that harm crops, or floods that threaten households. An important objective in predicting and mitigating natural disasters is risk assessment, which involves estimating marginal and joint probabilities of extreme events. On the other hand, data often contain little information about the rare events that lead to catastrophes, and therefore the methods employed must possess sufficient extrapolative power to adequately describe rare event probabilities. In addition, environmental data are typically non-stationary, which often requires incorporating external factors into models of extreme events.

Statistics of extremes has long established itself as an effective methodology for the estimation and analysis of extreme events. While the statistical theory of univariate extremes is, at present, essentially well developed and well understood, see, e.g., Beirlant et al. (2004), multivariate extreme value theory still involves a large number of challenges, see Hazra et al. (2025); Nolde and Wadsworth (2022); Engelke and Hitz (2020) for the recent developments in the field.

In general, modelling multivariate extremes involves first estimating the marginal distribution tails and then the dependence structure of multivariate extremes. At the same time, the estimation of univariate extremes for environmental data, and precipitation in particular, has its own specific characteristics. The literature offers no consensus regarding the most appropriate distribution for modelling the tails of precipitation data and the distributions of its annual maxima. Although the Gumbel distribution was originally the dominant model for the latter, later studies have indicated that heavy-tailed distributions with large tail indices may provide a better fit (Koutsoyiannis 2004; Serinaldi and Kilsby 2014). Another approach that is popular among practitioners is the so-called metastatistical approach (Marani and Ignaccolo 2015), which is non-asymptotic in nature and is based on the Weibull distribution.

This paper is devoted to describing our approach to solving the tasks of the EVA (2025) Conference Data Challenge on estimation of extreme precipitation, see for details Cooley et al. (2026). A key step in our solution is the transition to the analysis of extremes of univariate sequences, whereas the Data Challenge tasks were originally formulated in a multivariate setting. For this purpose, alongside classical methods of statistics of univariate extremes, we employ the recently proposed block quantiles method (Jakubowski and Rodionov 2026) and hope that this work will help promote this method among practitioners.

The remainder of the paper is organised as follows. In Section 2, we provide the description of the data and target quantities of the Data Challenge. Section 3 presents our methodology for the Data Challenge tasks. The paper ends with Section 4, discussing the obtained results.

2 Data description and target quantities

The main objective of the EVA 2025 Data Challenge was to predict extreme precipitation in a specific but a priori unknown region, based on data over a long time period,

Cooley et al. (2026). Although the data are essentially generated using the CESM2 Large Ensemble Community Project (LENS2) they mimic real precipitation data and, for example, exhibit seasonality, serial and spatial dependence.

From each of the 50 LENS2 simulation runs, precipitation data of length 165 years ($60225 = 165 \times 365$ days per series), corresponding to a 5×5 grid at the same (unspecified) locations, were extracted for the Data Challenge. Participants were provided with the data from 4 out of these 50 runs, while the true values of the quantities to be estimated are computed empirically by averaging over all 50 runs. The observations of the target variable were scaled, and their units were changed on "Leadbetters". The competition target quantities are specified as follows:

T1. To estimate the expected number of times all 25 locations exceed 1.7 Leadbetters.
T2. To estimate the expected number of times at least 6 of the 25 sites exceed 5.7 Leadbetters.
T3. To estimate the expected number of times at least 3 of the 25 sites exceed 5 Leadbetters for a run of at least two consecutive days.

For each target quantity, it was required to provide not only the point estimate, but also a central 95% confidence interval, that is, a total of nine target values had to be provided. Further details can be found in Cooley et al. (2026).

3 Methodology and results

During the Data Challenge, we experimented with some multivariate extreme value approaches; however, a number of challenges, including limited computational resources and instability of estimates, prevented us from adequately adapting these models to the Data Challenge tasks and, in particular, from constructing reliable confidence intervals for the target quantities.

Instead, we observed that the Data Challenge tasks can be addressed using exclusively methods of *univariate* extreme value analysis. Indeed, each task can be reformulated as a univariate problem:

1. to estimate **T1**, it is enough to estimate the expected number of days on which the minimum precipitation value across all 25 locations exceeds 1.7 Leadbetters;
2. to estimate **T2**, it is enough to estimate the expected number of days on which the 6^{th} largest order statistic across all 25 locations exceeds 5.7 Leadbetters;
3. to estimate **T3**, it suffices to estimate the expected number of days on which the 3^{rd} largest order statistic of 2-day moving minima across all 25 locations exceeds 5 Leadbetters.

To address these problems, we follow the unified approach combining the Peaks-over-Threshold (POT) approach to univariate extremes with the recently proposed block quantiles method (Jakubowski and Rodionov 2026), which is described later in Section 3.3. The POT approach considers modelling the distribution of exceedances over a high threshold u, i.e., $Y - u | Y > u$, via the Generalised Pareto (GP) distribution model with scale parameter $\sigma_u > 0$ and shape parameter $\xi \in \mathbb{R}$, which is called the extreme value index (EVI). The GP distribution with cumulative distribution function

(cdf)

$$F(y; u, \sigma_u, \xi) = 1 - \left(1 + \xi y/\sigma_u\right)_+^{-1/\xi}, \quad y > 0, \tag{1}$$

where $x_+ = \max(0, x)$, is a theoretically justified model for describing exceedances over a high threshold, u, see Davison and Smith (1990). If $\xi = 0$, then $F(y; u, \sigma_u, 0)$ in (1) should be understood as $1 - \exp(-y/\sigma_u)$, $y > 0$, that is the distribution of threshold exceedances is approximated by the exponential distribution. We also say in this case that the distribution of Y belongs to the Gumbel maximum domain of attraction (MDA), whereas if $\xi > 0$ we say that the distribution Y is heavy-tailed and belongs to the Fréchet MDA. The case $\xi < 0$ corresponds to the distribution of Y belonging to the Weibull MDA.

In Section 3.4, we reduce the estimation of the target quantities to estimating the probabilities that the annual maxima of certain time series exceed the levels specified above in this section. Although the standard approach for estimating such probabilities is the Gumbel block maxima method (Beirlant et al. 2004), we instead use a combination of two methods: the POT approach applied to a thinned sample, together with extremal index estimation, and the recently proposed block quantiles method (see Section 3.3 for details), in order to make the most efficient use of the information contained in the data. We split the data into winter and summer seasons (see Section 3.2), assuming that the data are stationary within each season. We then apply the aforementioned approaches separately to the data from each season, selecting the best approach via a leave-one-sample-out cross-validation procedure; see Section 3.4.

Extreme values of stationary sequences often occur in clusters, so the distribution of the maxima of such sequences differs from that under independence. To account for this effect, the extremal index is typically introduced to the analysis. Let $\{X_i\}_{i\geq 1}$ be a strictly stationary sequence with a marginal cdf F, and $M_N = \max\{X_1, \ldots, X_N\}$. If the extremal index $\theta \in (0, 1]$ of this sequence exists, then the following approximation holds for all x and N large enough

$$P(M_N \leq x) \approx (F(x))^{\theta N}.$$

Notice also that if the sequence $\{X_i\}_{i\geq 1}$ is i.i.d., then $\theta = 1$. The previous relation can be rewritten as

$$P(M_N > x) \approx 1 - \exp(-\theta\, N P(X_1 > x)),$$

see, e.g., Coles (2001), which we use to estimate the target quantities within the POT approach. In this study, we estimate θ using the intervals estimator (Ferro and Segers 2003), with the threshold selected via the automatic procedure proposed in Markovich and Rodionov (2024). In turn, we estimate the probability $P(X_1 > x)$ using maximum likelihood estimation under the GP model (1) (see Davison and Smith (1990)), with the threshold selected using a different procedure, namely that introduced in Murphy et al. (2025). To account for clustering of extremes, we apply maximum likelihood estimation to exceedances of a thinned sample, ensuring that the extremal index of the sample is greater than 0.9, see the details of thinning in Section 3.2. For the time series used to estimate the target quantities **T1** and **T2**, the thinning factor is $s = 2$, whereas we set $s = 3$ when estimating **T3**.

The choice of the block quantiles method alongside the POT approach is motivated by our belief that, for the data under study, precipitation extremes from October to April are better modelled using the Weibull-type model (this should not be confused with the Weibull extreme value distribution, which belongs to the Generalized Extreme Value (GEV) distribution family) rather than the GP distribution, please refer to Section 3.2 and the last paragraph in Appendix B of the Supplementary Material. We say that a cdf F is of Weibull type (Gardes et al. 2011; Goegebeur and Guillou 2010) if there exists $\lambda > 0$ such that

$$\lim_{x \to +\infty} \frac{\log(1 - F(tx))}{\log(1 - F(x))} = t^{1/\lambda}, \quad t > 0. \tag{2}$$

The parameter λ is called the Weibull tail index. This class of distributions belongs to the Gumbel (not Weibull!) maximum domain of attraction; representatives of this class include the normal, exponential, gamma and Weibull distributions.

Our approach to the Data Challenge tasks is close to that adopted by Davison and Rodionov (2026), who demonstrated that the Weibull-type model is comparable to, and in some cases outperforms, the Generalized Pareto distribution in terms of the quality of precipitation extremes estimation.

The rest of this section is organised as follows. In Section 3.1 we provide evidence of non-stationary nature of LENS2 data and distinguish two seasons, see also Appendix A in the Supplementary Material. Section 3.2 presents our approach to addressing revealed seasonality and model selection for precipitation extremes, see also Appendix B. In Section 3.3 we provide details of the recently proposed block quantiles method (Jakubowski and Rodionov 2026) for estimating maxima of stationary sequence. Finally, Section 3.4 is devoted to describing our approach to estimating the Data Challenge target quantities, which is based on a combination of the POT approach and the block quantiles method.

3.1 Exploratory data analysis

Precipitation data are often non-stationary, and the data under study are no exception. The right panel of Fig. 2 in Cooley et al. (2026) shows evidence of seasonality in the data, and we also discuss this effect in detail in Appendix A of the Supplementary Material. The analysis presented there indicates that the main source of non-stationarity is a pronounced annual cycle with wetter cold seasons and drier warm seasons, whereas no significant trend is observed in the data.

To investigate the seasonal structure in more detail, we present Fig. 1, which shows, for the locations (1,1), (2,2), (3,3) and for each month, box plots of precipitation values observed over the 165-year period that exceed the 95^{th} empirical quantile of all observations in that month. We use the data from the first available run only, the results obtained using the remaining runs and other locations are similar, see also Figure S3 in the Supplementary Material. Figure 1 shows that the distributions of precipitation extremes from May to September are quite similar, whereas both the sample medians and the variability of these values in the remaining "winter" months are substantially higher. This motivates us to consider two seasons in the given data,

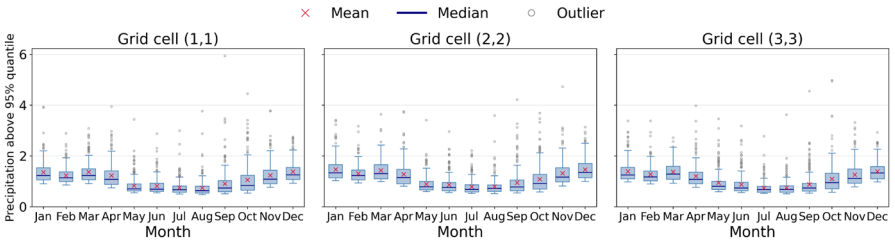


Fig. 1 Boxplots of precipitation values, which exceed the 95% sample quantile in the corresponding month, for locations (1,1), (2,2) and (3,3)

namely a "summer" season including observations from May to September, and a "winter" season including observations from October to April.

We also note that the largest precipitation values observed in the summer months are often higher than those of the winter months. This suggests that the distribution tails of precipitation in summer may be heavier than those in winter. The analysis presented in the next section confirms this conjecture. Therefore, seasonality cannot be treated as a purely additive or multiplicative component in a precipitation model for the given data, and in order to apply methods that assume stationarity of the observations, it is necessary at least to analyse each selected season separately.

3.2 Accounting for seasonality and model selection

To check the conjecture that the distribution tails of precipitation values in the Data Challenge are heavier in the summer months than in the winter months, we plot the maximum likelihood estimates of the extreme value index for the summer and winter seasons, as well as separately for each month, see Fig. 2. To mitigate the extremal dependence of threshold exceedances when applying the maximum likelihood method, we thin the data (i.e., we selected every sth observation from the sample of threshold exceedances) ensuring that the corresponding extremal index values exceed 0.9, which we estimate using the intervals approach (Ferro and Segers 2003). It turns out that taking $s = 2$ is sufficient for this purpose. The value 0.9 is chosen because, when the extremal index exceeds this threshold, the extremes of a stationary time series may be treated as independent. Hereafter we assume that observations from each individual month form a stationary sequence. For this analysis, we selected the locations with coordinates (1,1), (2,2), (3,3), (4,4), and (5,5), and computed the estimates using only the first run of the available data. The results for the remaining runs are similar.

As noted in the Introduction, precipitation extremes are commonly modelled using regularly varying distributions, with the extreme value index typically around 0.1 (Koutsoyiannis 2004; Serinaldi and Kilsby 2014). Figure 2 generally confirms this conclusion. Indeed, we observe that for all five locations selected for analysis, the EVI estimates approach 0.2 as k (the number of largest order statistics used for estimation) increases for the months of the summer season, while for the months of the winter season, the estimates for all five locations tend to be equal to or slightly below 0. This

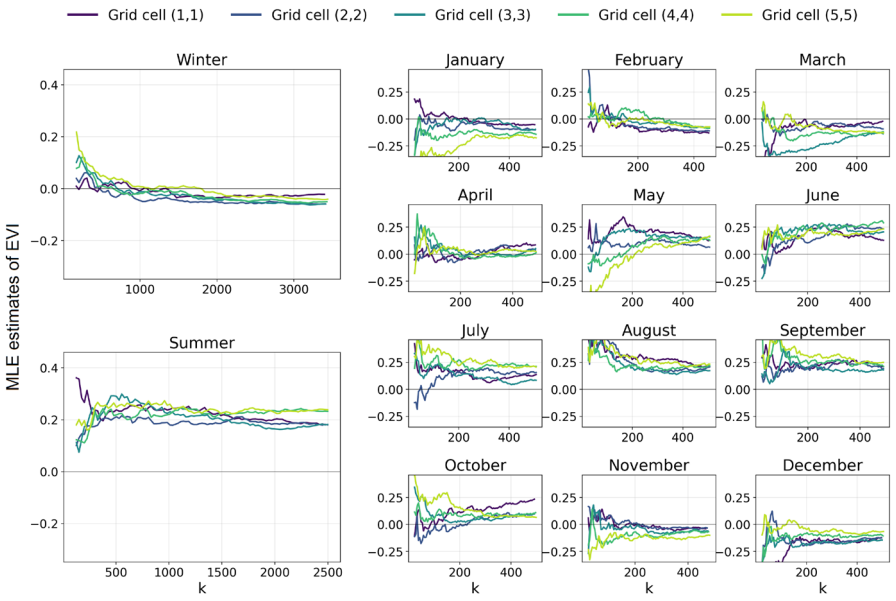


Fig. 2 Left panel: maximum likelihood estimates of the extreme value index for the winter (top) and summer (bottom) seasons as a function of k, the number of the largest order statistics used in the estimation. Right panel: corresponding estimates obtained using observations from each month

confirms our conjecture that the tail behaviour differs between the summer and winter seasons. In particular, precipitation distributions in the summer season exhibit heavy tails, whereas those in the winter season appear to belong to the Gumbel MDA and thus exhibit lighter tails.

The behaviour of EVI estimates for the observations in the winter season suggests that the appropriate model for precipitation extremes in this season might be the Weibull-type model (2). To check this, we employ the test proposed in Rodionov (2024) to select between the GPD model with a positive extreme value index (that is, having heavy tails, denote them by $\mathcal{RV}$) and the Weibull-type model ($\mathcal{W}$); for additional details, see Davison and Rodionov (2026). It is based on the following statistic

$$\widehat{R}_{k,n} = \log\frac{k}{n} - \frac{1}{k}\sum_{i=n-k+1}^{n} \log\left[1 - F_0\Big(q_0\big(\frac{n}{k}\big) + \frac{X_{i,n} - X_{n-k,n}}{X_{n-k,n} - X_{n-2k,n}}\Big(q_0\big(\frac{n}{k}\big) - q_0\big(\frac{n}{2k}\big)\Big)\Big)\right],$$

where $k \ll n$, F_0 is some cdf separating distributions of two above models in terms of tail heaviness, and $q_0(t) = \inf\{x : F_0(x) = 1 - 1/t\}$ is the tail quantile function of F_0. Then, to distinguish between $\mathcal{RV}$ and $\mathcal{W}$ at the significance level α, the following procedure is applied

$$\text{select} \begin{cases} \mathcal{RV}, & \text{if } \widehat{R}_{k,n} \geq 1 + \frac{u_\alpha \sigma}{\sqrt{k}}, \\ \mathcal{W}, & \text{otherwise,} \end{cases} \tag{3}$$

where u_α denotes the α-quantile of $N(0, 1)$, and $\sigma = \sqrt{1 + 1/\log^2 4}$. For the above procedure, we select

$$F_0(x) = 1 - \exp\big(-0.3\,(\log x)^2\big), \quad x > 1,$$

as in Davison and Rodionov (2026), and $\alpha = 0.05$. The tail of F_0 is heavier than those of Weibull-type distributions and lighter than regularly-varying tails, that is it is a suitable separating distribution function for these two distribution families, see for details Rodionov (2024).

Figure 3 shows the values of the statistic $\sqrt{k}(\hat{R}_{k,n} - 1)/\sigma$ with respect to k, the number of the largest order statistics used for testing, based on thinned precipitation data from locations (1,1), (2,2), (3,3), and (4,4) in the winter and summer seasons with the thinning factor $s = 2$. For the winter season, the statistic values are below the critical level $u_{0.05}$ in all cases except for small k, indicating evidence against the null hypothesis and favouring the Weibull-type model. In contrast, for the summer season, the statistic values fluctuate around the critical level, which does not provide definitive evidence in favour of either hypothesis; thus, there are no sufficient grounds to reject the null hypothesis of a regularly varying tail behaviour.

The above analysis indicates that, for precipitation data from all locations during the winter season, the Weibull-type model is appropriate, whereas the classical GPD model can be used for modelling precipitation in the summer season. This conclusion is also supported by the behaviour of the quantile-quantile plots for precipitation extremes, which can be found in Appendix B of the Supplementary Material. It is worth noting that results from different locations appear similar, which suggests that

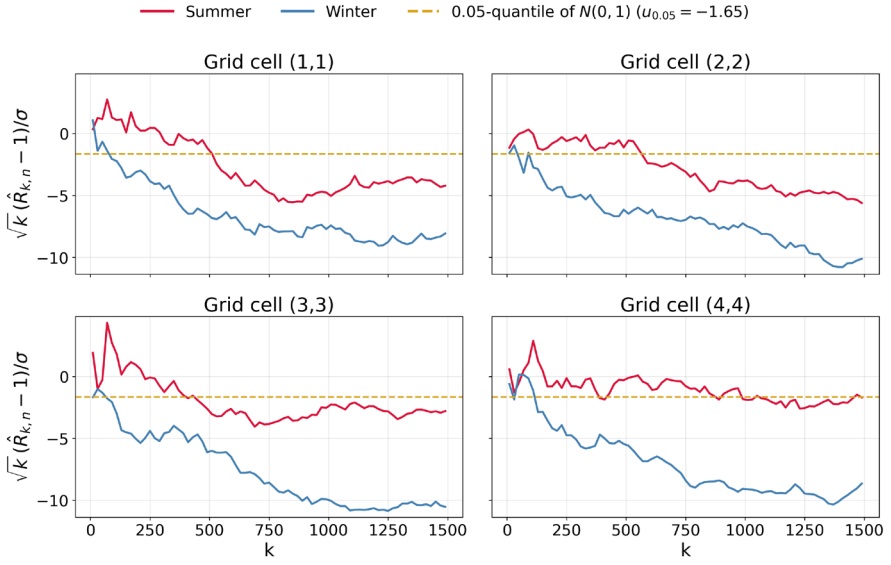


Fig. 3 The values of the statistic $\sqrt{k}(\hat{R}_{k,n} - 1)/\sigma$ as a function of k for thinned precipitation data from locations (1,1), (2,2), (3,3), and (4,4) in the winter (blue line) and summer (red line) seasons. The yellow line corresponds to the 0.05-quantile of $N(0, 1)$, the thinning factor is $s = 2$

the extreme value indices for summer seasons are consistent across locations, as well as the Weibull indices for winter seasons.

However, further analysis will be conducted on time series composed of the order statistics from the vector of observations across all 25 available locations. In particular, we use the minima and the 6$^{\text{th}}$ largest order statistics over all locations to estimate the target quantities **T1** and **T2**, respectively, whereas to estimate **T3**, we use the 3$^{\text{rd}}$ largest order statistics of 2-day moving minima over all locations. Procedures analogous to those carried out above were applied to these time series, and the results were very similar to those obtained for the original sequences and thus were omitted. The only notable difference is that we set the thinning factor $s = 3$ for the third series, while keeping $s = 2$ for the first two, which can be explained by the fact that the series is constructed using 2-day moving minima. Thus, we assume hereafter that these random variables are also regularly varying in the summer season and of Weibull type in the winter season.

3.3 Block quantiles method

As mentioned above, to estimate rare event probabilities as required by the Data Challenge, we apply the POT approach and the recently proposed block quantiles method. This section is devoted to a description of the latter.

The block quantiles is actually a method for estimating the tail of a so-called phantom distribution function (originally introduced in O'Brien (1987)) of a stationary sequence, which generalises the notion of the extremal index. We say that a stationary sequence $\{X_i\}_{i\geq 1}$ admits a phantom distribution function if for some cdf G

$$\sup_{x\in\mathbb{R}} |P(M_N \leq x) - G^N(x)| \to 0, \quad N \to \infty, \tag{4}$$

where $M_N = \max(X_1, \ldots, X_N)$. The existence of a phantom distribution function for a stationary sequence is more general property than the existence of the extremal index, Jakubowski (1993). The relation (4) implies that if the tail of the phantom distribution function is estimated, this will also provide an estimate for the distribution of the maximum, given sufficiently large N. Notice that if G can be selected as F^θ in Eq. 4, where F is the marginal cdf of the sequence $\{X_i\}_{i\geq 1}$, then θ is the extremal index of this sequence.

Let $q_{G^N}(\gamma)$ and $q_{M_N}(\gamma)$ denote the γ-quantiles of G^N and a cdf of M_N, respectively, where $\gamma \in (0, 1)$. To estimate the quantiles of M_N, we use the approximation $q_{M_N}(\gamma) \approx q_{G^N}(\gamma) = q_G(\gamma^{1/N})$ following immediately from Eq. 4. In turn, the block quantiles method assumes selection of a parametric model for G to estimate its extreme quantiles. The GP distribution may be used for this purpose, although alternative tail models, such as the Weibull-type distribution, may also be considered.

The idea of the block quantiles method is close to the classical Gumbel block maxima method, see, e.g., Beirlant et al. (2004). One should divide an available sample on several (say ℓ) blocks of consecutive observations, find the largest observation in every block (denote them by $m_1, \ldots, m_\ell$), and compare the empirical quantiles of generated "sample" with the quantiles of G^r of the corresponding levels, where

$r = \lfloor n/\ell \rfloor$ and n is the sample size. Let κ be the vector of parameters of G. Then to estimate κ, one may solve the regression problem

$$m_{i,\ell} = q_{G^r(\kappa)}(i/(\ell+1)) + \varepsilon_i$$

for some range of i values, where $m_{1,\ell} \le \ldots \le m_{\ell,\ell}$ denote the order statistics of the "sample" $m_1, \ldots, m_\ell$, and $\{\varepsilon_i\}_{i=1}^{\ell}$ denote i.i.d. errors with mean zero. The key difference between the block quantiles and Gumbel block maxima methods is that in the former we merge data obtained from various block sizes and apply regression techniques already to the aggregated dataset. Below, we provide the algorithm of the block quantiles method.

Algorithm 1.

1. Select an interval $[\alpha_{\min}, \alpha_{\max}]$ of possible values of quantile levels (usually $\alpha_{\min} = 0.01$, $\alpha_{\max} = 0.99$).
2. Select a set of block sizes $\{r_1, \ldots, r_s\}$.
3. For a fixed block size r, divide n observations on $\ell = \lfloor n/r \rfloor$ consecutive blocks of size r, where $\lfloor \cdot \rfloor$ denotes a floor function.
4. Find the largest observation in every block and denote them by $m_1, \ldots, m_\ell$.
5. Form an array $\mathbf{A}_r$ consisting of all pairs

$$(y, x) = \left(m_{i,\ell}, \left(\frac{i}{\ell+1} \right)^{1/r} \right)$$

   for $i \in \{1, \ldots, \ell\}$ with $\alpha_{\min} \le i/(\ell+1) \le \alpha_{\max}$.
6. Repeat steps 3-5 for each $r \in \{r_1, \ldots, r_s\}$ and derive the arrays $(\mathbf{A}_{r_1}, \ldots, \mathbf{A}_{r_s})$. Obtain the array $\mathbf{A}$ of pairs (y_i, x_i) by merging $\mathbf{A}_{r_1}, \ldots, \mathbf{A}_{r_\ell}$.
7. For the chosen model of G, find the estimate $\hat{\kappa}$ by solving the corresponding regression problem

$$y = q_{G(\kappa)}(x) + \varepsilon, \quad (y, x) \in \mathbf{A} \tag{5}$$

   with respect to κ.

We apply to y in Eq. 5 the transformation $y_i' = y_i - \min_i y_i + c$, where $(y_i, x_i) \in \mathbf{A}$ and c is defined in Eq. 6, which improves the quality of estimation. In this study, we use

$$s = \lfloor \sqrt{2n^{1/2}} \rfloor, \text{ and } r_i = n^{\frac{1}{2} + \frac{i-1}{2s}}, \quad i = 1, \ldots, s.$$

This choice is motivated by that the parameter estimates obtained via the block quantile method are consistent as $r_i \to \infty$ and $n/r_i \to \infty$, $i = 1, \ldots, s$; see Jakubowski and Rodionov (2026). It also covers a wide range of different block-size asymptotic regimes. We also note that the estimates obtained using the block quantiles method are fairly robust to the choice of r_1, which constitutes an advantage of the method. The minimum block size r_1 plays a role somewhat similar to that of the threshold in the POT approach: to a large extent, the value of the former determines how many distinct observations from the original sample $\{X_i\}_{i=1}^{n}$ enter the array $\mathbf{A}$ as block maxima. However, unlike the POT approach applied to i.i.d. observations, not all observations

exceeding a given threshold are used for parameter estimation, moreover, the selected observations may be used multiple times in the estimation procedure.

How to solve the regression equation (5) merits a separate discussion. For this, we use three methods: standard least squares, Theil–Sen (Sen 1968) and minimizing the L_p loss function (in this study, we set $p = 1$)

$$L_p(\kappa) = \sum_{(x,y)\in\mathbf{A}} |y - q_{G(\kappa)}(x)|^p$$

by ADAM optimizer (Kingma and Ba 2015). The first two methods are applicable only if the regression problem (5) can be linearised, i.e. transformed to the form

$$f(y) = \kappa_1 + \kappa_2 g(x) + \epsilon$$

for some f and g, where ϵ denotes an error. It is possible if, in turn, G can be represented as

$$G(x) = 1 - h((x/b)^a), \quad x > c, \tag{6}$$

for some h, a, b, and c, which holds, in particular, for Pareto and Weibull distributions. To improve the stability of the block quantiles method for estimating the quantiles of M_N, we average three quantile estimates obtained using the least squares, Theil–Sen, and L_p-ADAM variants of the method.[1] We refer to this modification as the "consensus" approach. However, this complicates the construction of confidence intervals for the block quantiles method. Finally, to estimate the probability that M_N exceeds the level u, the equation $\hat{Q}(p) = u$ is solved with respect to p, where $\hat{Q}$ is the quantile function of M_N estimated by the consensus variant of the block quantiles method.

3.4 Estimation of target quantities

According to preliminary estimates, in all three Data Challenge tasks the expected number of days over 165 years on which the events of interest occur is less than one; therefore, it is natural to assume that such extreme events occur no more than once per year with probability close to one. Consequently, rather than analysing daily probabilities of exceeding a high level in tasks **T1–T3**, we may consider the probability that the annual (or seasonal) maximum exceeds the specified level and then multiply this probability by 165, corresponding to the number of years in the observational record. This observation applies to the application of both the block quantiles method and the POT approach, which makes the required analysis simpler.

To derive these probabilities, we estimate the annual exceedance probabilities separately for each of the four runs and then average them, in accordance with the procedure used to obtain the true values of the target quantities in the Data Challenge. These

[1] It is worth explaining why we adopt an averaging approach rather than standard model selection and choosing the best of the three methods. In extensive simulations conducted by one of the authors in Jakubowski and Rodionov (2026), it was observed that, in practice, two scenarios most commonly occur: either the least squares and Theil–Sen methods underestimate the quantile value while the L_p-ADAM method overestimates it, or vice versa. The reasons for this effect are not yet fully understood. However, averaging the predictions of these three methods yields, in the vast majority of cases, more accurate results than any of the individual methods taken separately.

annual exceedance probabilities are, in turn, obtained from the corresponding probabilities estimated separately for the winter and summer seasons as follows

$$p_a = 1 - (1 - p_w)(1 - p_s), \tag{7}$$

where p_a, p_w, and p_s denote the annual, winter, and summer exceedance probabilities, respectively. This is in agreement with the $\mathbf{B}_T$ condition (Doukhan et al. 2015), that is necessary for a stationary sequence to admit a phantom distribution function, which we assume to exist. We estimate these seasonal probabilities using samples formed by concatenating the 165 blocks of observations for the corresponding season and then applying the POT approach, in combination with extremal index estimation (see the beginning of Section 3 for its description), as well as the block quantiles method (see Section 3.3) to these two merged samples, setting $N = 153$ and $N = 212$ for the summer and winter seasons, respectively.

To select the best model for estimating the target rare event probabilities for each of the six stationary sequences (three target quantities $\times$ two seasons), we apply the following procedure. Since for each season and target quantity we have data from four independent runs, we employ a cross-validation procedure. In each iteration, we combine three runs to form a training set of samples, while the remaining run is used as a test (validation) sample, so it is actually leave-one-series-out cross-validation across the four independent runs. As a quality metric, we use the MAE

$$\mathrm{MAE}(u) = \tfrac{1}{4} \textstyle\sum_{i=1}^{4} \left| \tilde{p}_{\setminus i}(u) - \hat{p}_i(u) \right|$$

between $\tilde{p}_{\setminus i}(u)$, $1 \le i \le 4$, the averages of three estimates of the probabilities of exceeding level u, each constructed from a distinct run in the training set using the algorithm described above (i.e., one run is not used at this step), and the corresponding probability estimated empirically from the remaining validation run, $\hat{p}_i(u)$, $1 \le i \le 4$, where u ranges over values from the 0.995 to the 0.999-quantile of the empirical distribution of daily precipitation in the corresponding season. For comparison, we use the POT approach and the consensus variant of the block quantiles method with the Weibull distribution chosen as G (denoted below as BQ (Weibull)). We also applied the block quantiles method with the Pareto distribution chosen as G; however, in all cases, this method was outperformed by at least one of the two previous methods.

Table 1 shows the top-performing methods according to cross-validation for the winter and summer seasons, along with the final estimates of the target quantities obtained via a combination of these methods and corresponding confidence intervals.

It remains for us to explain how we build confidence intervals for the target quantities. For this, we use the following bootstrap method. Since the data in both seasons to

Table 1 Final estimates of the expected number of extreme events

Target	Summer	Winter	Target estimate	Confidence bounds
T1	POT	BQ (Weibull)	0.1013	[0.0548; 0.3318]
T2	POT	BQ (Weibull)	0.1934	[0.0820; 0.3918]
T3	POT	POT	0.0139	[0.0013; 0.0329]

which we apply our methods are assumed to form a stationary sequence, the standard bootstrap destroys the temporal dependence structure of the data. Hence, it is natural to partition the data into reasonable blocks and to generate bootstrap samples from these blocks rather than from individual observations. In our analysis, years serve as the natural blocks. The bootstrap procedure is therefore as follows: we first draw a bootstrap sample $(i_1, \ldots, i_{165})$ from the uniform distribution on the set $\{1, \ldots, 165\}$, and then concatenate the blocks corresponding to the years with indices $(i_1, \ldots, i_{165})$ into a single sample, while preserving the order of the indices. In this way, we construct 400 bootstrap samples and, for each of them, apply the selected algorithm described earlier in this section to obtain an estimate of the probability of exceeding the specified level. The number 400 is chosen due to limitations in the computational resources available to the members of our team during the Data Challenge.[2] Finally, given a sample of 400 bootstrap-based probability estimates, we construct a quantile bootstrap confidence interval of a level 0.95, that is, taking the empirical quantiles at levels 0.025 and 0.975 of this sample as the lower and upper bounds. A similar approach employed to construct bootstrap confidence bounds for extreme quantiles demonstrated performance comparable to empirical confidence intervals, see Jakubowski and Rodionov (2026); thus, we expect the method to work properly for estimating rare event probabilities as well.

4 Discussion

In this section, we briefly describe our approach to solving the tasks of the EVA (2025) Conference Data Challenge and explain the rationale behind our choice of methods. Despite the multivariate nature of the tasks, they can be reformulated as univariate problems, which we exploited in our solution. In essence, our approach to each of the three Data Challenge tasks was the same. After detecting nonstationarity in the data, we split them into two seasons, summer (from May to September) and winter (from October to April), within which the observations appeared to form stationary sequences. This was supported by exploratory data analysis as well as an examination of monthly extreme value indices. Moreover, our analysis indicated that the data in the winter season do not exhibit regular variation, which led us to consider a different model for the winter season than the standard Generalised Pareto distribution (namely, the Weibull-type model) in order to better capture the behaviour of extremes during this period.

Note that in the Data Challenge tasks, the goal was to estimate the expected number of days over 165 years on which a quantity, depending on daily precipitation observations at 25 locations, exceeds a certain high threshold. Accounting for dependence in the data would make this potentially difficult to estimate. However, preliminary estimates, including empirical ones, indicated that these quantities are so small (less

[2] For the first preliminary target quantity, a central bootstrap confidence interval was constructed based on 1,000 bootstrap samples. As it was nearly identical to the interval obtained using 400 samples, we opted to use the latter number in the subsequent construction of confidence intervals.

than one) that, with probability close to one, the events of interest occur no more than once per year. This allowed us to shift from considering daily exceedances to annual exceedances. Thus, to estimate the target quantities of the Data Challenge, it was sufficient to multiply the estimated probability that the annual maximum exceeds the specified level by 165.

To estimate the probability that the annual maximum exceeds a given level, we split the data into summer and winter seasons, estimated the exceedance probabilities separately for each season, and then combined them using (7). For each season, we selected from three methods: the classical POT approach combined with the extremal index estimation, and the recently proposed block quantiles method (Jakubowski and Rodionov 2026) with Pareto and Weibull distributions. As a result, the POT approach was identified as the best method for the summer season based on cross-validation, whereas in two out of three Data Challenge tasks, the block quantiles method with the Weibull distribution was the best for the winter season. Finally, confidence intervals for the target quantities were obtained using the bootstrap method.

We acknowledge that the optimal solution to the Data Challenge tasks would likely involve the use of methods of multivariate extreme value analysis rather than the univariate analysis we employed. However, by the time we were able to fit multivariate extreme value models, we no longer had the opportunity to assess their performance through the estimation of preliminary target quantities. As a result, we focused our solution on univariate methods. Moreover, having found that the distribution of extreme precipitation in the winter season is well described by a Weibull-type model, we supposed that using the block quantiles method, performing well for distributions of this type, could provide us a competitive advantage.

The statistical theory of extreme values is currently regarded as classical and well-understood in the univariate case and to a large degree for univariate time series. However, we believe that discoveries are still possible in this field — at least in cases where the distribution tails are not regularly varying. In this study, we aimed to demonstrate the performance of the recently proposed block quantiles method for the analysis of environmental extremes. We place considerable hope in this method and believe that it has the potential to exert a strong impact on extreme value statistics, not only in its univariate part but also in multivariate analysis, which often involves the estimation of marginal distribution tails.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s10687-026-00549-4.

Author Contributions The analysis described in the paper was performed by R. Guryev, V. Kuzmenko, and I. Saralpov. S.Platinskii contributed to solving Data Challenge tasks at the initial stage and suggested switching to univariate analysis. I. Rodionov wrote the main manuscript text, and all authors reviewed the manuscript.

Funding The authors declare that no funds, grants, or other support were received during the preparation of this manuscript.

Data Availability The manuscript uses the data available at this link https://eva2025.unc.edu/data-challenge/ to solve the EVA (2025) Conference Data Challenge tasks.

Declarations

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

Beirlant, J., Goegebeur, Y., Segers, J., Teugels, J.: Statistics of Extremes. Theory and Applications. Wiley, London (2004)

CESM2 Large Ensemble Community Project (LENS2). https://www.cesm.ucar.edu/community-projects/lens2

Coles, S.: An Introduction to Statistical Modeling of Extreme Values, Springer (2001)

Cooley, D., Hector, E., Shaby, B., Wadsworth, J.: Editorial: EVA 2025 conference data challenge. Extremes (2026)

Davison, A.C., Rodionov, I.: Predicting extreme precipitation using block quantiles, (2026). (Working paper.)

Davison, A.C., Smith, R.L.: Models for exceedances over high thresholds (with discussion). J. Roy. Stat. Soc. B **52**(3), 393–442 (1990)

Doukhan, P., Jakubowski, A., Lang, G.: Phantom distribution functions for some stationary sequences. Extremes **18**, 697–725 (2015)

Engelke, S., Hitz, A.S.: Graphical models for extremes (with discussion). J. R. Stat. Soc. Ser. B Stat Methodol. **82**, 871–932 (2020)

Ferro, C.A.T., Segers, J.: Inference for clusters of extreme values. J. Roy. Stat. Soc. B **65**, 545–556 (2003)

Gardes, L., Girard, S., Guillou, A.: Weibull tail-distributions revisited: a new look at some tail estimators. Journal of Statistical Planning and Inference **141**(1), 429–444 (2011)

Goegebeur, J., Guillou, A.: Goodness-of-fit testing for Weibull-type behavior. Journal of Statistical Planning and Inference **140**(6), 1417–1436 (2010)

Hazra, A., Huser, R., Bolin, D.: Efficient modeling of spatial extremes over large geographical domains. J. Comput. Graph. Stat. **34**(3), 795–811 (2025)

Jakubowski, A.: An asymptotic independent representation in limit theorems for maxima of nonstationary random sequences. Ann. Probab. **21**, 819–830 (1993)

Jakubowski, A., Rodionov, I.: Block quantiles: the method for estimation of maxima of stationary sequence, (2026). (Working paper.)

Kingma, D.P., Ba, J.: ADAM: A method for stochastic optimization. International Conference on Learning Representations (ICLR-2015), vol. 13, pp. (2015)

Koutsoyiannis, D.: Statistics of extremes and estimation of extreme rainfall: II. Empirical investigation of long rainfall records. Hydrol. Sci. J. **49**, 591–610 (2004)

Marani, M., Ignaccolo, M.: A metastatistical approach to rainfall extremes. Adv. Water Resour. **79**, 121–126 (2015)

Markovich, N.M., Rodionov, I.V.: Threshold selection for extremal index estimation. Journal of Nonparametric Statistics **36**(3), 527–546 (2024)

Murphy, C., Tawn, J.A., Varty, Z.: Automated threshold selection and associated inference uncertainty for univariate extremes. Technometrics **67**(2), 215–224 (2025)

Nolde, N., Wadsworth, J.L.: Linking representations for multivariate extremes via a limit set. Adv. Appl. Probab. **54**(3), 688–717 (2022)

O'Brien, G.L.: Extreme values for stationary and Markov sequences. Ann. Probab. **15**, 281–292 (1987)

Rodionov, I.: Location- and scale-free procedures for distinguishing between distribution tail models. Electronic Journal of Statistics **18**(2), 5330–5369 (2024)
Sen, P.K.: Estimates of the regression coefficient based on Kendall's tau. J. Am. Stat. Assoc. **63**(324), 1379–1389 (1968)
Serinaldi, F., Kilsby, C.G.: Rainfall extremes: Toward reconciliation after the battle of distributions. Water Resour. Res. **50**, 336–352 (2014)

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.