



**HAL**  
open science

## Generalized random forest for extreme quantile regression

Lucien M. Vidagbandji, Alexandre Berred, Cyrille Bertelle, Laurent Amanton

► **To cite this version:**

Lucien M. Vidagbandji, Alexandre Berred, Cyrille Bertelle, Laurent Amanton. Generalized random forest for extreme quantile regression. *Communications in Statistics - Simulation and Computation*, 2025, pp.1-24. <10.1080/03610918.2025.2543854>. <hal-05314096>

**HAL Id: hal-05314096**

**<https://hal.science/hal-05314096v1>**

Submitted on 14 Oct 2025

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire HAL, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



HAL Authorization

# Generalized random forest for extreme quantile regression

Lucien M. Vidagbandji<sup>1\*</sup>, Alexandre Berred<sup>1</sup>, Cyrille Bertelle<sup>2</sup>,  
Laurent Amanton<sup>2</sup>

<sup>1</sup>Université Le Havre Normandie, Normandie Univ., LMAH UR 3821,  
76600, Le Havre, France.

<sup>2</sup>Université Le Havre Normandie, Normandie Univ., LITIS, 76600, Le  
Havre, France.

\*Corresponding author(s). E-mail(s):

[mahutin-lucien.vidagbandji@univ-lehavre.fr](mailto:mahutin-lucien.vidagbandji@univ-lehavre.fr);

Contributing authors: [alexandre.berred@univ-lehavre.fr](mailto:alexandre.berred@univ-lehavre.fr);  
[cyrille.bertelle@univ-lehavre.fr](mailto:cyrille.bertelle@univ-lehavre.fr); [laurent.amanton@univ-lehavre.fr](mailto:laurent.amanton@univ-lehavre.fr);

## Abstract

Quantile regression is a statistical method which, unlike classical regression, aims to predict the conditional quantiles. Classical quantile regression methods face difficulties, particularly when the quantile under consideration is extreme, due to the limited number of data available in the tail of the distribution, or when the quantile function is complex. We propose an extreme quantile regression method based on extreme value theory and statistical learning to overcome these difficulties. Following the Block Maxima approach of extreme value theory, we approximate the conditional distribution of block maxima by the generalized extreme value distribution, with covariate-dependent parameters. These parameters are estimated using a method based on generalized random forests. Applications on simulated data show that our proposed method effectively addresses the mentioned quantile regression issues and highlights its performance compared to other quantile regression approaches based on statistical learning methods. We apply our methodology to daily meteorological data from the Fort Collins station in Colorado (USA).

**Keywords:** Extreme quantile regression, generalized extreme value distribution, generalized random forest, maximum likelihood estimator, block maxima method.

# 1 Introduction

Extreme events, defined by their scarcity and potential for severe impact, frequently dominate the news and capture our imagination due to their unpredictability and the losses they cause. Examples of such extreme events include floods, earthquakes, rogue waves, financial crises, to name just a few. Understanding or anticipating the occurrence of these events allows for a deeper analysis of the associated risks. Extreme quantile regression is an appropriate tool in this regard. It is used to model the relationship between explanatory variables (covariates) and the extreme conditional quantiles of a response variable. This approach helps in understanding how factors influence the extreme values of a distribution. Extreme events occur in the tails of distributions, so for risk analysis, it's mainly a question of estimating extreme quantiles, i.e. quantiles at high level  $\tau$ . Specifically, in the univariate context, if  $Y \in \mathcal{Y} \subset \mathbb{R}$  represents a random variable describing a risk factor dependent on a set of covariates represented by the random vector  $X \in \mathcal{X} \subset \mathbb{R}^p$ , the objective is to estimate the extreme conditional quantile defined by:

$$\mathcal{Q}_x(\tau) = \inf\{y : F_{Y|X=x}(y) \geq \tau\}, \quad x \in \mathcal{X}, \quad (1)$$

with  $\tau$  close to 1 and  $F_{Y|X=x}$  being the conditional distribution of  $Y|X = x$  (see [Koenker and Bassett \(1978\)](#)). Observations in the tails of the distribution are often rare and pose a particular challenge when estimating the quantity (1). This difficulty arises from the limited nature of available data in these extreme regions of the distribution. Indeed, rare events are often crucial in various fields such as risk management, strategic planning, and decision-making. In the context of extreme quantile regression, it becomes imperative to develop robust and accurate methods to handle these infrequent observations, in order to improve the reliability of models and predictions in situations where extreme outcomes play a significant role. This is the case, for example, in risk analysis, in hydrology when estimating a water level that has not yet been observed ([Gumbel \(1963\)](#), [Hu and Franzke \(2020\)](#)).

Let  $n$  denote the sample size available for analysis and  $\tau = \tau_n$  (dependent on the sample size) the level of the quantile we seek to estimate. Classical quantile estimation methods work well when  $n(1 - \tau_n) \rightarrow \infty$  as  $\tau_n \rightarrow 1$  (for  $n \rightarrow +\infty$ ). In this case, the quantile we seek to estimate is within the sample, and there is also a large amount of data in the region of the quantile to be estimated. However, the situation differs when  $n(1 - \tau_n) \in [0, +\infty[$  as  $\tau_n \rightarrow 1$  (when  $n \rightarrow +\infty$ ). In this latter case, estimation involves extrapolating beyond the data range or into the distribution's tail, and  $\tau_n$  is referred to as an extreme quantile level. In other words, the sought quantile lies outside the range of the available sample, making estimation more complex and requiring specific approaches to handle these borderline situation. Thanks to the asymptotic results of extreme value theory ([de Haan and Ferreira \(2006\)](#)), extrapolation beyond this range is possible. Based on this theory, [Chernozhukov \(2005\)](#) proposed an extreme quantile regression method within the framework of a linear relationship between the response variable  $Y$  and the predictor variable  $X \in \mathbb{R}^p$ . Several other authors have proposed extreme quantile regression models, including [Yu and Moyeed \(2001\)](#), [Koenker and Hallock \(2001\)](#), [Daouia et al. \(2013\)](#), [Takeuchi et al. \(2006\)](#), [Laksaci and Maref \(2009\)](#),

Gardes and Stupfler (2015), Gardes and Stupfler (2019), and El Methni and Stupfler (2017).

The second difficulty faced by classical extreme quantile regression methods arises when the dimension of the predictor space  $\mathcal{X}$  is large (i.e.,  $p$  is large) and the relationship between the characteristic variables and the response variable is complex. In high dimensions, a simple quantile regression model may introduce additional bias (Gnecco et al. (2024)). Recent literature has seen the development of machine learning approaches to address this challenge. First, we can cite the work of Meinshausen and Ridgeway (2006), who proposed a method based on the random forest method of Breiman (2001). Athey et al. (2019) proposed a quantile regression model based on the generalized random forest method. Their methods encounter difficulties mainly when the quantile of interest is extreme (as shown by the simulation studies in section 4). Quantile regression methods that combine extreme value theory and machine learning have emerged very recently and are competitive with existing methods when the quantile of interest is extreme and when the predictor space is high-dimensional. To our knowledge, the first model proposed in this sense is that of Farkas et al. (2024), based on regression trees and the Peaks Over Threshold (POT) approach of extreme value theory. Next, we can mention the method of Velthoen et al. (2023), which models the distribution of equation (1) using the conditional generalized Pareto distribution and the gradient boosting method. Pasche and Engelke (2024) propose a method that still combines extreme value theory to approximate the conditional distribution of equation (1) and neural networks to estimate the parameters of this distribution. Gnecco et al. (2024) propose a method combining the POT approach and generalized random forests.

Previous work has mainly relied on the *Peaks Over Threshold* approach to address the problems associated with estimating extremes. Although this method is widely recognized for its effectiveness, it nevertheless presents several notable limitations, particularly when the only available data consist of block maxima aggregated over fixed periods — such as annual maxima derived from long historical time series or large-scale simulations. In such contexts, the block maxima (BM) approach offers a more suitable alternative. This method is extensively used in environmental sciences, where the Generalized Extreme Value (GEV) distribution has proven particularly well-suited for modeling phenomena such as annual or monthly maximum temperatures, or extreme river discharges. As highlighted by Ferreira and De Haan (2015), the BM approach presents several practical advantages. It is particularly appropriate when only periodic maxima data are available. Moreover, it allows for a certain degree of temporal dependence within blocks — for instance, in the presence of seasonality or short-term dependence — which is advantageous in situations where the assumption of independent and identically distributed observations is difficult to justify. Finally, the BM approach is often simpler to implement, as block periods naturally align with existing temporal structures (e.g., years or seasons). In this work, we focus on addressing the problem of extreme quantile regression using the block maxima approach. This method constitutes a relevant alternative to POT-based method in many applied contexts, particularly when data availability or temporal structure constraints hinder the use of threshold exceedance methods. The generalized extreme value distribution

is given by:

$$G_{(\mu,\sigma,\xi)}(z) = \begin{cases} \exp\left(-\left(1 + \xi \frac{z - \mu}{\sigma}\right)_+^{-\frac{1}{\xi}}\right) & \text{if } \xi \neq 0, \\ \exp\left(-\exp\left(-\frac{z - \mu}{\sigma}\right)\right) & \text{if } \xi = 0, \end{cases} \quad (2)$$

defined on  $\{z : 1 + \xi \frac{z - \mu}{\sigma} > 0\}$  where  $\mu \in \mathbb{R}$ ,  $\sigma > 0$ , and  $\xi \in \mathbb{R}$  are respectively the location, scale, and shape parameters, with  $a_+ = \max\{0, a\}$ ,  $\forall a \in \mathbb{R}$ .

In this study, applying the block maxima approach, we approximate the conditional distribution  $Y|X = x$  of block maxima by a generalized extreme value distribution, whose parameters are functions of the features  $x \in \mathcal{X}$ . These parameters are estimated by a weighted maximum likelihood method, with weights obtained using the generalized random forest method developed by [Athey et al. \(2019\)](#). Section 2 provides a reviews of quantile regression, the BM approach to extreme value theory and the generalized random forest method. Section 3 presents our proposed model in detail. Section 4 examines the results obtained when applying our extreme quantile regression method to simulated data under several scenarios. Finally, section 5 presents an application to daily weather data from the Fort Collins station in Colorado (USA).

## 2 Background

### 2.1 Generalized extreme value distribution

In this section, we will present the approach of extreme value theory that we will use to address the first quantile regression challenge mentioned in the introduction: the block maxima approach. This method is based on the results of [Fisher and Tippett \(1928\)](#) and [Gnedenko \(1943\)](#), which provide the possible asymptotic distributions of the maximum of a sequence of random variables  $X_1, \dots, X_m$  drawn independently and identically distributed from a variable  $X$  with probability distribution  $F$ . These authors showed that there exist normalization constants  $a_m > 0$  and  $b_m \in \mathbb{R}$  such that:

$$\lim_{m \rightarrow +\infty} F^m(a_m x + b_m) = G_\xi(x), \quad x \in \mathbb{R} \text{ and } \xi \in \mathbb{R}, \quad (3)$$

where  $G_\xi$  is a non-degenerate probability distribution defined by:

$$G_\xi(x) = \exp\left(-\left(1 + \xi x\right)^{-\frac{1}{\xi}}\right) \text{ with } 1 + \xi x > 0,$$

which is called the extreme value distribution. Any function  $F$  that satisfies equation (3) is said to belong to the max-domain of attraction of the extreme value distribution  $G_\xi$  and is denoted as  $F \in \mathcal{D}(G_\xi)$  ([de Haan and Ferreira \(2006\)](#)). Considering  $Y_1, \dots, Y_N$  as a sequence of independent and identically distributed random variables according to the random variable  $Y$  with distribution function  $F$ , the block maxima method involves dividing the data into  $n$  blocks of equal size  $m > 1$  (or nearly equal), denoted as

$$B_{k,m} = \{Y_{(k-1)m+1}, \dots, Y_{km}\}, \text{ with } k = 1, \dots, n.$$

For any  $m > 1$ , the distribution of  $Z_k = \max_{B_{k,m}}(Y_i)$  is  $F^m$  and satisfies (3) with a certain normalization constant  $a_m > 0$  and  $b_m \in \mathbb{R}$ . Therefore, its distribution is given by the generalized extreme value distribution with parameters  $(a_m, b_m, \xi_0)$ . The BM method assumes that these block maxima, denoted  $Z_k$ , exactly follow the GEV distribution (explicitly defined by (2)), and that the sequence of random variables  $Z_1, \dots, Z_n$  thus formed is also independent and identically distributed. The specificity of this method lies in the choice of the optimal block size. Increasing the block size  $m$  leads to an increase in the variance of the estimation, while decreasing the block size introduces a bias. Therefore, it is essential to find a trade-off between bias and variance when defining the blocks to ensure accurate estimation. The BM method is commonly presented, discussed, and employed in the literature for modeling extremes. Among the reference works, one can refer to [Coles \(2001\)](#) and [de Haan and Ferreira \(2006\)](#). The general form of the Generalized Extreme Value distribution, introduced by [Jenkinson \(1955\)](#), is given by the function  $G_{(\mu, \sigma, \xi)}(\cdot)$  defined in equation (2). Depending on the sign of  $\xi$ , three domains of attraction are distinguished: the Fréchet domain of attraction if  $\xi > 0$ , the Gumbel domain of attraction if  $\xi = 0$ , and the Weibull domain of attraction if  $\xi < 0$ . The  $\tau$ -th extreme quantile of the Generalized Extreme Value distribution is obtained using equation (1) and is given by:

$$Q_\tau = \begin{cases} \mu + \frac{\sigma}{\xi} \left( (-\ln(\tau))^{-\xi} - 1 \right) & \text{if } \xi \neq 0, \\ \mu - \sigma \ln(-\ln(\tau)) & \text{if } \xi = 0, \end{cases}$$

where  $\tau$  close to 1,  $\mu \in \mathbb{R}$ ,  $\sigma > 0$ , and  $\xi \in \mathbb{R}$ . Estimating the quantile of the GEV distribution therefore requires the estimation of the parameters  $\mu$ ,  $\sigma$  and  $\xi$ . Among the various methods available, the maximum likelihood method is the most commonly used and is the focus of this study. The theoretical validity of this method for GEV parameter estimation has been demonstrated by [Dombry \(2015\)](#); [Dombry and Ferreira \(2019\)](#) and [Bücher and Segers \(2017\)](#). Denoting the parameter vector of GEV as  $\theta = (\mu, \sigma, \xi)$ , the negative log-likelihood of the GEV distribution for the sample  $z_1, \dots, z_n$  is given by:

$$L(\theta; z_1, \dots, z_n) = \frac{1}{n} \sum_{i=1}^n \ell_\theta(z_i), \quad (4)$$

where  $\ell_\theta$  is defined by:

$$\ell_\theta(z_i) = \begin{cases} \log(\sigma) + \left(1 + \frac{1}{\xi}\right) \log\left(1 + \xi \frac{z_i - \mu}{\sigma}\right) + \left(1 + \xi \frac{z_i - \mu}{\sigma}\right)^{-\frac{1}{\xi}} & \text{if } 1 + \xi \frac{z_i - \mu}{\sigma} > 0, \\ +\infty & \text{otherwise,} \end{cases}$$

when  $\xi \neq 0$ , and

$$\ell_\theta(z_i) = \log(\sigma) + \left(\frac{z_i - \mu}{\sigma}\right) + \exp\left(-\frac{z_i - \mu}{\sigma}\right),$$

when  $\xi = 0$ . The maximum likelihood estimator  $\hat{\theta}$  satisfies

$$\hat{\theta} \in \arg \min_{\theta \in \Theta} L(\theta; z_1, \dots, z_n),$$

where  $\Theta \subset \mathbb{R} \times ]0, +\infty[ \times \mathbb{R}$ .

## 2.2 Generalized random forests

In this section, we introduce the machine learning method we will be using to address the second problem, which aims to handle the high-dimensional space  $\mathcal{X}$  of predictors and capture the complex relationship between the response variable and the features. The generalized random forests (Athey et al. (2019)) is a generalization of the classical random forest method proposed by Breiman (2001). It retains the attractive features of classical random forests but allows customization of the loss function used for tree construction. Random forest is a learning method used for both classification and regression tasks. It belongs to the family of ensemble learning methods and provides a non-parametric estimation of the conditional mean. It involves aggregating  $B$  trees trained in parallel on bootstrap samples from the original training set, similar to the bagging method introduced by Breiman (1996). A key difference with bagging lies in how feature variables are used to split nodes in each tree. Instead of using all features at every node split, a random subset of features is selected. This random selection introduces additional variability between the trees, promoting diversity and improving the model's generalization (Breiman (2001)). Each tree provides an estimate of the conditional mean, which is obtained as the function minimizing the mean squared error. If  $(X_i, Y_i) \in \mathcal{X} \times \mathcal{Y}$  where  $\mathcal{X} \subset \mathbb{R}^p$  and  $\mathcal{Y} \subset \mathbb{R}$ , random forests are used to estimate  $\mu(x) = \mathbb{E}(Y_i | X_i = x)$ . Let  $\eta_b(x)$  denote the predicted value by the  $b^{\text{th}}$  tree for data  $x \in \mathcal{X}$ . In the case of regression, this prediction is given by:

$$\eta_b(x) = \sum_{i=1}^n \frac{\mathbb{1}_{\{X_i \in R_b(x)\}} Y_i}{|\{i : X_i \in R_b(x)\}|}, \quad b = 1, \dots, B,$$

where  $R_b(x) \subset \mathcal{X}$  denotes the region containing  $x$  in tree  $b$ , and  $|E|$  denotes the cardinal of  $E$ . The random forest gives as its prediction the average of the predictions of the  $B$  trees, thus it predicts the value  $\eta(x)$  given below for data  $x$ :

$$\begin{aligned} \eta(x) &= \frac{1}{B} \sum_{b=1}^B \eta_b(x) \\ &= \sum_{i=1}^n w_n(x, X_i) Y_i, \end{aligned}$$

with

$$w_n(x, X_i) = \frac{1}{B} \sum_{b=1}^B \frac{\mathbb{1}_{\{X_i \in R_b(x)\}}}{|\{i : X_i \in R_b(x)\}|}. \quad (5)$$

The  $w_n(x, X_i)$ , for  $i = 1, \dots, n$ , represent the similarity weights assigned by the random forest for each observation  $X_i$  given  $x \in \mathcal{X}$ .

While the similarity weights  $w_n(x, X_i)$  assigned by the generalized random forests (GRF) are computed using the same formula as in the classical forest, the key

difference lies in the loss functions employed during tree construction. In classical random forests, a high weight is typically assigned to an observation  $X_i$  when  $\mathbb{E}(Y|X = X_i) \approx \mathbb{E}(Y|X = x)$  (Meinshausen and Ridgeway (2006)). However, as shown by Athey et al. (2019) (Figure 2) and Gnecco et al. (2024) (Figure 1), these weights can still be large even when  $\mathcal{Q}(Y|X = X_i) \not\approx \mathcal{Q}(Y|X = x)$ , indicating that classical forests often fail to capture heterogeneity in the conditional quantile function. This limitation is addressed in GRF by incorporating the quantile loss into the tree construction process, enabling the method to target conditional quantiles directly. This approach has been successfully applied for both moderate and extreme quantile modeling by Athey et al. (2019) and Gnecco et al. (2024), respectively. Forests-based methods have the advantage of requiring few tuning parameters to provide effective predictions. This characteristic makes generalized random forests a natural choice for constructing similarity weights, especially as they capture the heterogeneity of conditional quantiles, a central aspect in our study. Unlike classical approaches such as kernel or nearest-neighbor methods, which are often limited when faced with non-linear structures or high-dimensional data, GRF exploit tree structure to accurately model local variability. They thus produce similarity weights that are both more robust and better suited to estimating extreme quantiles. What’s more, compared with techniques such as gradient boosting or neural networks, GRF combine simplicity of use with sound theoretical foundations (Athey et al. (2019)). They also offer better high-dimensional adaptability than generalized additive models (Koenker (2011)). Furthermore, the BM approach is better suited to random forests, as it generates i.i.d. observations from fixed block maxima, in line with the assumptions underlying random forests. Conversely, the POT method produces potentially correlated excesses, requiring additional adjustments, and relies on an often empirical threshold, which can weaken learning. This combination of robustness and flexibility makes GRF-based weights particularly well-suited to our framework. For clarity, we denote  $w_i(x) \equiv w_n(x, X_i)$  for all  $x \in \mathcal{X}$  and  $i = 1, \dots, n$ .

### 2.3 Quantile regression

In classical regression, given a random vector  $(X, Y)$ , we seek to predict the conditional mean  $E[Y|X = x]$ , which is a quantity that summarizes the behavior of  $Y$  given  $X = x$  around the center of distribution. However, applications in risk assessment require knowledge of the tail behavior of its distribution. Quantile regression, introduced by Koenker and Bassett (1978), addresses this by providing a richer description than classical regression. The latter focuses on the estimation of the conditional quantile defined in (1), which in the case of a continuous conditional distribution is given by:

$$\mathcal{Q}_\tau(Y|X = x) = F_{Y|X=x}^{-1}(\tau), \quad x \in \mathcal{X} \text{ and } \tau \in (0, 1). \quad (6)$$

In the following, we denote  $\mathcal{Q}_x(\tau)$  to represent the conditional quantile function  $\mathcal{Q}_\tau(Y|X = x)$ . To the best of our knowledge, the first appearance of the random forest method in quantile regression, addressing our second problem stated in the introduction, dates from the work of Meinshausen and Ridgeway (2006). Considering

$(X_1, Y_1), \dots, (X_n, Y_n)$  as a sequence of independent and identically distributed random vectors from  $(X, Y) \in \mathcal{X} \times \mathcal{Y}$ , the author estimates the conditional distribution  $F(\cdot|X = x)$  by:

$$\hat{F}(y|X = x) = \sum_{i=1}^n w_n(x, X_i) \mathbb{1}_{\{Y_i \leq y\}}, \quad y \in \mathbb{R} \text{ and } x \in \mathcal{X},$$

where the weights  $w_n(x, X_i)$  are obtained using the random forest method by [Breiman \(2001\)](#) as defined in [\(5\)](#). Another method of quantile regression is proposed by [Athey et al. \(2019\)](#) using the generalized random forest method described in [section \(2.2\)](#), with the quantile loss function

$$\rho_\tau(c) = c(\tau - \mathbb{1}_{\{c < 0\}}).$$

Although these methods address high-dimensional problems and capture complex relationships between the dependent variable and covariates, they encounter difficulties when the order of the quantile is extreme. Additional research has therefore been conducted to address this issue by combining extreme value theory, mainly the POT approach, and statistical learning methods. Among these efforts, notable contributions include [Farkas et al. \(2024\)](#), [Gnecco et al. \(2024\)](#), [Velthoen et al. \(2023\)](#), and [Pasche and Engelke \(2024\)](#). The quantile regression method we propose and detail in the next section builds upon these recent developments.

### 3 GEV Extremal Random Forest

In this section, we present our method of extreme quantile regression that addresses the issues outlined in the introduction. To address the first issue, we model the tail of the conditional distribution  $F(\cdot|X = x)$  from [equation \(6\)](#) using the generalized extreme value distribution, as described in [section \(2.1\)](#). To tackle the second issue, we employ the generalized random forest method to determine the weights  $w_n(x, X_i)$ , which will be used to estimate the parameters of the GEV distribution. These parameters are estimated using weighted maximum likelihood estimators, following the approach proposed by [Gnecco et al. \(2024\)](#) within the peaks-over-threshold framework. To proceed, let's consider an independent and identically distributed sample  $(X_1, Y_1), \dots, (X_N, Y_N)$  from the random vector  $(X, Y)$ . Assume  $N = n \times m$  and divide the sample into  $n$  blocks of size  $m$ , such that the  $k^{th}$  block for  $k = 1, \dots, n$  is given by:

$$B_{k,m} = \{(X_{(k-1)m+1}, Y_{(k-1)m+1}), \dots, (X_{km}, Y_{km})\}.$$

Noting  $Z_{k,m} = \max\{Y_i : (X_i, Y_i) \in B_{k,m}\}$  and  $X_{k,m}$  the  $X_i$  corresponding to the  $Y_i$  maximizing  $\{Y_i : (X_i, Y_i) \in B_{k,m}\}$ . Thus, for all  $m > 1$ ,  $\{(X_{k,m}, Z_{k,m})\}_{k=1, \dots, n}$  represents the sample of block maxima relative to the variable  $Y$ . For any  $x \in \mathcal{X}$ , we assume that the distribution  $F_x$  of  $Z_{k,m}|X_{k,m} = x$  follows a Generalized Extreme Value distribution with parameter depending to the covariate  $x$ . More precisely,  $\mu(\cdot) : \mathcal{X} \rightarrow \mathbb{R}$  is the location parameter function,  $\sigma(\cdot) : \mathcal{X} \rightarrow \mathbb{R}_+^*$  is the scale parameter function

and  $\xi(\cdot) : \mathcal{X} \rightarrow \mathbb{R}$  is the shape parameter function. The conditional GEV distribution is obtained by substituting  $\theta = (\mu, \sigma, \xi)$  in (2) with  $\theta(x) = (\mu(x), \sigma(x), \xi(x))$  for all  $x \in \mathcal{X}$ . The  $\tau$ -th extreme quantile of the conditional GEV is obtained using equation (6) and is given for all  $x \in \mathcal{X}$  by:

$$Q_x(\tau) = \begin{cases} \mu(x) + \frac{\sigma(x)}{\xi(x)} \left( (-\ln(\tau))^{-\xi(x)} - 1 \right) & \text{if } \xi(x) \neq 0, \\ \mu(x) - \sigma(x) \ln(-\ln(\tau)) & \text{if } \xi(x) = 0. \end{cases} \quad (7)$$

Estimating this conditional quantile involves estimating the parameters  $\mu(x)$ ,  $\sigma(x)$ , and  $\xi(x)$  for all  $x \in \mathcal{X}$ . Our proposed method here consists of estimating these parameters and then substituting them into equation (7) to obtain the estimation of  $Q_x(\tau)$ . For this purpose, we propose a weighted form of the maximum likelihood estimator given in (4). As proposed by [Gnecco et al. \(2024\)](#) for the POT approach, we estimate  $\theta(x)$  by  $\hat{\theta}(x) = (\hat{\mu}(x), \hat{\sigma}(x), \hat{\xi}(x))$ , which is a weighted maximum likelihood estimator, i.e., it minimizes  $L_n(\theta; x)$  defined by:

$$L_n(\theta; x) = \sum_{i=1}^n w_i(x) \ell_{\theta}(z_i) \text{ for all } x \in \mathcal{X}, \quad (8)$$

with  $\ell_{\theta}(z_i)$  defined in section (2.1). To capture the heterogeneity of the quantile function and improve estimation in a high-dimensional feature space, we obtain the weights  $w_i(x)$  for  $i = 1, \dots, n$  using the generalized random forests method, as described in section (2.2).

It should be noted that the support of the GEV distribution depends on the extreme value index  $\xi$ , which is unknown (see [Dombry \(2015\)](#)). The usual regularity conditions ensuring good asymptotic properties of the maximum likelihood estimator (MLE) are not satisfied. This problem is first studied by [Smith \(1985\)](#), who shows that there is no maximum likelihood estimator if  $\xi \leq -1$ , that asymptotic normality holds if  $\xi > -0.5$ , and consistency holds if  $\xi > -1$ . [Dombry \(2015\)](#) demonstrated the existence and consistency of this estimator for the GEV distribution parameters when  $\xi > -1$ . The asymptotic normality of the MLE for the block maxima approach was proven by [Dombry and Ferreira \(2019\)](#). In general,  $L_n(\theta; x)$  does not admit a global minimum, and we refer to the work of [Dombry \(2015\)](#) to state that  $\hat{\theta}(x)$  is a weighted maximum likelihood estimator if  $\hat{\theta}(x)$  is a local minimum. Thus, we define  $\hat{\theta}(x)$  to be the vector that minimizes  $L_n(\theta; x)$  over a large compact set  $\Theta \subset \mathbb{R} \times ]0, +\infty[ \times ]-1, +\infty[$ . The weighted maximum likelihood estimator is given by:

$$\hat{\theta}(x) \in \arg \min_{\theta \in \Theta} L_n(\theta; x). \quad (9)$$

We take the intermediate order  $\tau_0 = 0.8$  as considered in the works of [Gnecco et al. \(2024\)](#) and [Velthoen et al. \(2023\)](#). Thus, we estimate the conditional quantile of order

$\tau > \tau_0$  by:

$$\hat{Q}_x(\tau) = \begin{cases} \hat{\mu}(x) + \frac{\hat{\sigma}(x)}{\hat{\xi}(x)} \left( (-\ln(\tau))^{-\hat{\xi}(x)} - 1 \right) & \text{if } \hat{\xi}(x) \neq 0, \\ \hat{\mu}(x) - \hat{\sigma}(x) \ln(-\ln(\tau)) & \text{if } \hat{\xi}(x) = 0, \end{cases} \quad (10)$$

where  $\hat{\mu}(x)$ ,  $\hat{\sigma}(x)$  and  $\hat{\xi}(x)$  are the estimated parameters obtained using the weighted maximum likelihood method. The algorithm (1) details our method for predicting the extreme conditional quantiles, which we have named GEV-erf. This method combines the use of the generalized extreme value distribution with generalized random forests to estimate extreme conditional quantiles precisely and efficiently.

Among the parameters of the GEV distribution, the shape parameter  $\xi$  is of particular importance, as it provides information on the shape of the tail of the distribution. A distribution is considered to have a heavy tail when  $\xi > 0$ , a light tail when  $\xi = 0$ , and a finite endpoint distribution (i.e.  $x_F = \{y \in \mathbb{R}, F(y) < 1\}$  finite) when  $\xi < 0$ . Because of its importance, this parameter has received considerable attention in the literature. Many methods have been proposed for estimating  $\xi$ , the most notable examples being Hill's estimator, which focuses on the most important data points for estimating tail behavior, and Pickands' estimator. In line with the proposals of Gnecco et al. (2024) for maximum likelihood in the context of the POT approach, and Bücher et al. (2021) for the maximum likelihood of the GEV, we penalize the negative log-likelihood in the following form:

$$L_n^{pen}(\theta, x) = \sum_{i=1}^n w_i(x) \ell_{\theta}(z_i) + \lambda (\xi - \xi_0)^2.$$

The parameter  $\xi_0$  used here corresponds to the shape parameter derived from the unconditional GEV distribution, which means it is obtained by setting  $w_i(x) = 1$  for all indices  $i \in \{1, \dots, n\}$  in equation (9). The penalty parameter  $\lambda$  is determined by the cross-validation method, thus regularizing the negative log-likelihood in the estimation process. The resulting estimator,  $\hat{\theta}_{pen}(x)$ , is the weighted and penalized maximum likelihood estimator, which adjusts the GEV parameters by considering both the data and the model complexity, thereby ensuring better generalization. It is defined as :

$$\hat{\theta}_{pen}(x) \in \arg \min_{\theta \in \Theta} L_n^{pen}(\theta, x).$$

The algorithm to obtain the conditional quantile using  $\hat{\theta}_{pen}(x)$  is similar to the algorithm (1) with the difference that the log-likelihood  $L_n(\theta, x)$  is replaced by the penalized likelihood  $L_n^{pen}(\theta, x)$ . In the simulation study described in section (4), we provide a detailed explanation of the validation method used to determine the parameter  $\lambda$ . This method includes a series of tests and analyses to ensure the accuracy and reliability of the estimates. Specifically, we discuss the different steps of the validation process, the criteria used to evaluate the performance of the method, and the adjustments made to optimize the estimation of the parameter  $\lambda$ . The algorithm detailed below outlines our proposed method.

## Algorithm

The algorithm for our extreme quantile regression method, denoted GEV-erf, consists of two sub-algorithms: GEV-erf-fit and GEV-erf-predict. The first sub-algorithm, GEV-erf-fit, is used for training the generalized random forest to obtain the weights  $w_i(x)$  and to obtain the sample formed by the block maxima. The second sub-algorithm, GEV-erf-predict, is used to make predictions of the GEV distribution parameters and hence extreme conditional quantiles. It is presented as follows :

---

**Algorithm 1** GEV-erf

---

Let  $\mathcal{D}_N = \{(X_i, Y_i)\}_{i=1\dots N}$  denote the original sample,  $\mathcal{D}'_n = \{(X_i, Z_i)\}_{i=1\dots n}$  denote the sample of block maxima, where  $m$  is the block size, and  $\alpha$  is the vector containing hyperparameters associated with the generalized random forest.

- 1: **procedure** **GEV-erf-fit**
- 2: **Input:**  $(\mathcal{D}_N, m, \alpha)$
- 3:  $\mathcal{D}'_n \leftarrow \text{makeBloc}(\mathcal{D}_N, m)$
- 4:  $w_i(\cdot) \leftarrow \text{GRF}(\mathcal{D}'_n, \alpha)$
- 5: **Output:** GEV-erf  $\leftarrow (\mathcal{D}'_n, w_i(\cdot), m)$

- 1: **procedure** **GEV-erf-predict**
- 2: **Input:**  $(\text{GEV-erf}, x, \tau)$
- 3:  $\hat{\theta}(x) \leftarrow \arg \min_{\theta \in \Theta} L_n(\theta; x)$
- 4:  $\hat{Q}_x(\tau) \leftarrow \text{GEV}(\hat{\theta}(x))$
- 5: **Output:**  $\hat{Q}_x(\tau)$

- The function **makeBloc** divides the initial sample into blocks of a given size  $m$  and returns the sample of block maxima.
  - The **GRF** function is used to adjust the weights using the generalized random forests method proposed by [Athey et al. \(2019\)](#).
- 

## 4 Simulation

In this section, we present a simulation study to demonstrate the effectiveness of our proposed extreme quantile regression method. We generate an independent and identically distributed sample of size  $N = 90,000$  from the random variable  $X$ , following a uniform distribution over  $[-1, 1]^p$ . We assume that the conditional distribution of

$Y|X = x$  is of the form  $\gamma(x)T_{\nu(x)}$ , where  $T_k$  denotes the Student’s distribution with  $k$  degrees of freedom. We conduct various experiments by varying the values of  $\gamma(x)$  and  $\nu(x)$ . The choice of block size is a key challenge in the BM approach, especially since, as shown in Appendix C, the GEV-erf model is sensitive to this parameter. We therefore select a fixed block size of  $m = 40$  for all scenarios, as our analysis (see Appendix C) indicates that model performance remains stable for block sizes ranging from 25 to 50.

In our first simulation study (scenario 1), we set the dimension of the feature space to  $p = 40$ , and evaluate the performance of our method. We then compare its ability to address the first concern stated in the introduction with other methods based on statistical learning approaches. The  $N$  data points are used for the training process, while the evaluation of the method is conducted on an independent sample from the training set,  $\{(x_i, y_i)\}_{i=1}^{n'}$ , where  $x_i \in \mathcal{X}$  for  $i = 1, \dots, n'$  is generated using the Halton sequence method by Halton (1964) with  $n' = 8000$ . In the second simulation study (scenario 2), we demonstrate our method’s capability to handle the second concern, namely complex dependence on covariates. For this, we conduct simulations with functions  $\gamma(\cdot)$  and  $\nu(\cdot)$  chosen to introduce complex dependence with the covariates. In the last simulation study of this section (scenario 3), we show the effectiveness of our method in addressing both concerns simultaneously.

We proceed with comparing the performance of our method with two other random forest-based approaches: Quantile Regression Forests (QRF) by Meinshausen and Ridgeway (2006) and Generalized Random Forests (GRF) by Athey et al. (2019). To assess these performances, we first employ the Mean Integrated Squared Error (MISE), which is computed as the average of  $m'$  Integrated Squared Error (ISE) values for the estimated conditional quantile, derived from repeating the training and testing process  $m'$  times. This metric aligns with the approach used by Gnecco et al. (2024) in his work. Additionally, we evaluate performance using two other metrics: the Mean Absolute Error (MAE) and the Median Absolute Error (MedAE). The ISE on the test sample  $\{(x_i, y_i)\}_{i=1}^{n'}$  is defined as follows:

$$ISE = \frac{1}{n'} \sum_{i=1}^{n'} \left( \hat{Q}_{x_i}(\tau) - Q_{x_i}(\tau) \right)^2, \quad (11)$$

where  $x \mapsto Q_x(\tau)$  is the true quantile function,  $x \in \mathcal{X}$  and  $\tau > 0.8$ .

## 4.1 Simulations scenarios

We demonstrate the capability of our proposed method across three scenarios:

▷ **Scenario 1:** We take

$$\gamma(x) = 1 + \mathbb{1}_{\{x_1 > 0\}} \text{ and } \nu(x) = 4 - (x_1^2 - 2x_2^2 + x_3^2).$$

The goal is to assess our method’s ability to estimate extreme conditional quantiles in the presence of noise and when  $p$  is high.

▷ **Scenario 2:** We take

$$\gamma(x) = 2 + \frac{1}{1 + \exp(x_1^2 + x_2)} \text{ and } \nu(x) = 4 - (x_1^2 - 2x_2^2 + x_3^2).$$

The objective is to evaluate our method's capability to predict extreme conditional quantiles under complex forms of the quantile function. We set the feature size to  $p = 40$ , where only the first three components contribute to the quantile function formation, and the others are noise.

▷ **Scenario 3:** We consider

$$\gamma(x) = 1 + 2\pi\varphi(2x_1, 2x_2) \text{ and } \nu(x) = 3 + \frac{7}{1 + \exp(4x_1 + 1.2)}.$$

In this study,  $\varphi$  represents the density of the centered bivariate normal distribution with unit variance and a correlation coefficient of 0.75. The parameters chosen here are the same as those used in experiment 3 of the simulation conducted by [Gnecco et al. \(2024\)](#). The aim of this scenario is to evaluate our method's ability to accurately estimate the conditional quantile in situations where the quantile function is complex, the dimension of feature variables is high, and noise is present.

## 4.2 Parameters tuning choice

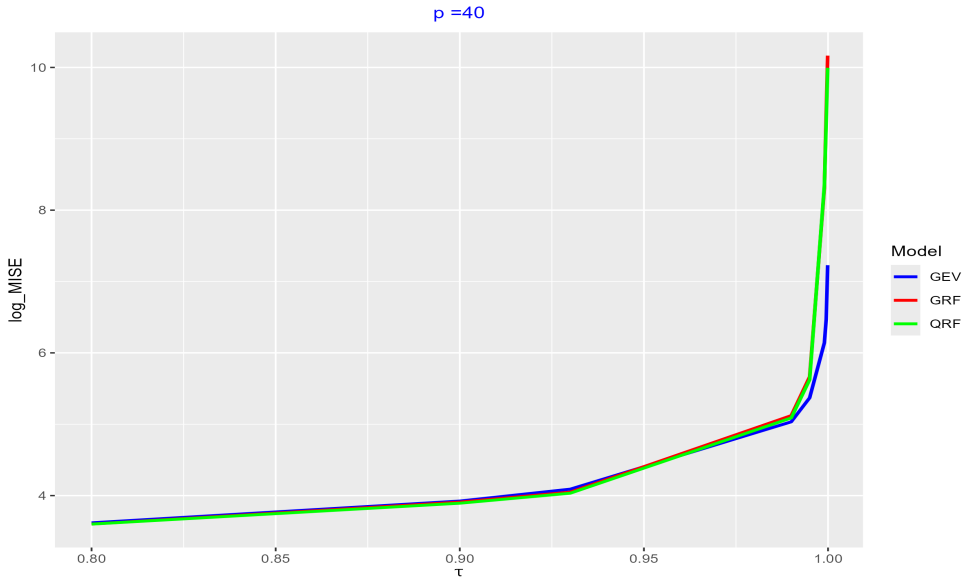
We obtain the penalty parameter  $\lambda$  and the parameters for generalized random forest (specifically `min.node.size` and `num.trees`) using cross-validation. Instead of using the classic form (as explained in Section 7.10 of [Hastie \(2017\)](#)), we follow the approach used by [Gnecco et al. \(2024\)](#), using GEV deviance as the metric. More clearly, if we consider that the training sample we have is  $\mathcal{D}_n = \{(x_i, z_i)\}_{i=1, \dots, n}$ , and we wish to perform a  $K$ -cross validation, the method involves partitioning the sample into  $K$  folds of approximately equal size, denoted as  $\mathcal{D}^j$  for  $j = 1, \dots, K$ . For each tuning parameter  $\alpha_1, \dots, \alpha_S$ , the model is trained on  $\mathcal{D}_n \setminus \mathcal{D}^j$  using the GEV-erf-fit algorithm, and then parameters  $\hat{\theta}(x)$  are estimated on  $\mathcal{D}^j$  for each  $j = 1, \dots, K$ . The performance metric, here the negative log-likelihood of GEV, is calculated for each fold, and the average error across the  $K$  folds provides an overall estimate of model performance. The  $\alpha_s$  achieving the best performance is selected, corresponding to minimizing the cross-validation error (CV-error) across  $s$ ,

$$CV(\alpha_s) = \frac{1}{K} \sum_{j=1}^K \sum_{(x_i, z_i) \in \mathcal{D}^j} \ell_{(\hat{\theta}(x_i), \alpha_s)}(z_i), \quad (12)$$

where  $\ell_{(\hat{\theta}(x), \alpha_s)}$  represents the negative log-likelihood defined in (4) and the  $\alpha_s$  for all  $s \in \{1, \dots, S\}$  designate the tuning parameters from which we seek to select the optimal choice. The selection of the parameters  $\lambda$  and `min.node.size` for the different scenarios is detailed in section (A) of the appendix. Our model is called GEV-erf, but we use the abbreviation "gev" to represent it in the various plots and tables showcasing the model performances.

### 4.3 Performance of GEV-erf with Scenario 1

In Scenario 1, we rigorously evaluated the performance of our method by comparing it with Quantile Regression Forest (QRF) proposed by [Meinshausen and Ridgeway \(2006\)](#) and Generalized Random Forest (GRF) by [Athey et al. \(2019\)](#). For this comparative analysis, we fixed the number of predictors to  $p = 40$  and plotted  $\log(\text{MISE})$  against the quantile level  $\tau$ . The MISE was computed from  $m' = 50$  repetitions of the ISE (defined in (11)), ensuring robust and statistically significant evaluation. Figure 1 clearly illustrates that our model is competitive compared to QRF and GRF models, especially for moderate quantile orders. This figure also highlights that our method provides better estimation of conditional quantiles, particularly as the quantile level is closed to 1.



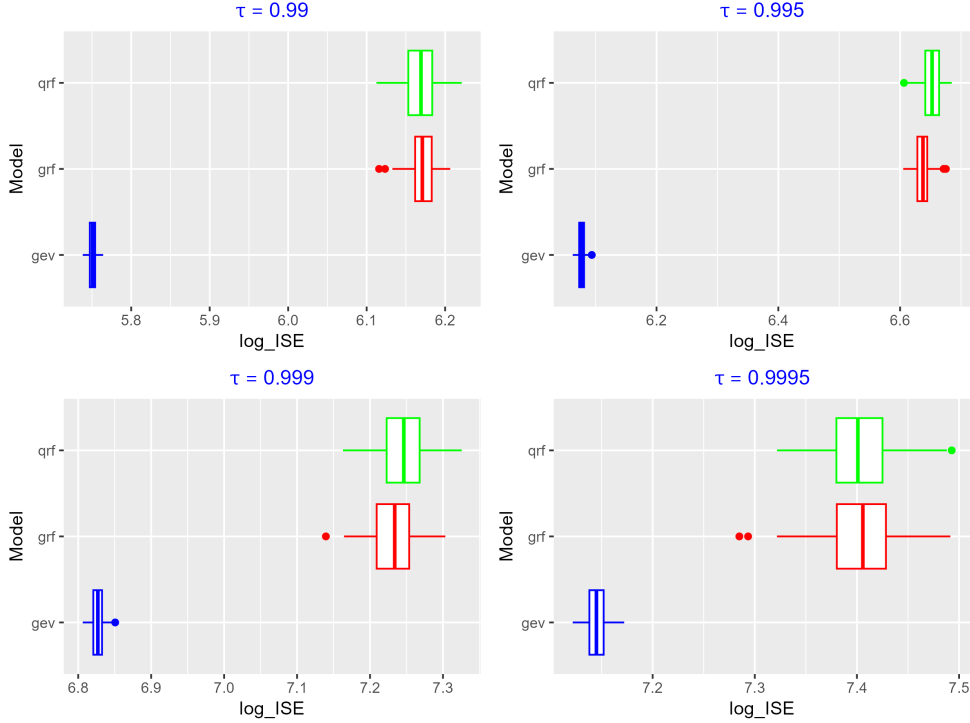
**Figure 1** Logarithm of MISE for different methods as a function of  $\tau$ .

Furthermore, our method demonstrates increased stability and reduced variance in quantile predictions, thanks to our innovative approach of local weighting. These results suggest that our model not only competes with existing methods but also outperforms them in scenarios where precise estimates of extreme quantiles are crucial. This is particularly relevant in applications requiring accurate risk management and decision-making based on reliable predictions of rare events.

### 4.4 Performance of GEV-erf with Scenario 2

In the figure 2, we evaluate the performance of our model for scenario 2. To do this, we present box plots of  $\log(\text{ISE})$  for different quantile orders  $\tau \in \{0.99, 0.995, 0.999, 0.9995\}$ , with  $p$  fixed at 40. The functions  $\gamma(\cdot)$  and  $\nu(\cdot)$

are chosen to make the conditional quantile function complex. The boxplots in this figure demonstrate that our model outperforms the QRF model and the GRF model for various quantile orders. This indicates the ability of our model to make accurate predictions even when the quantile function is complex. This scenario showcases our



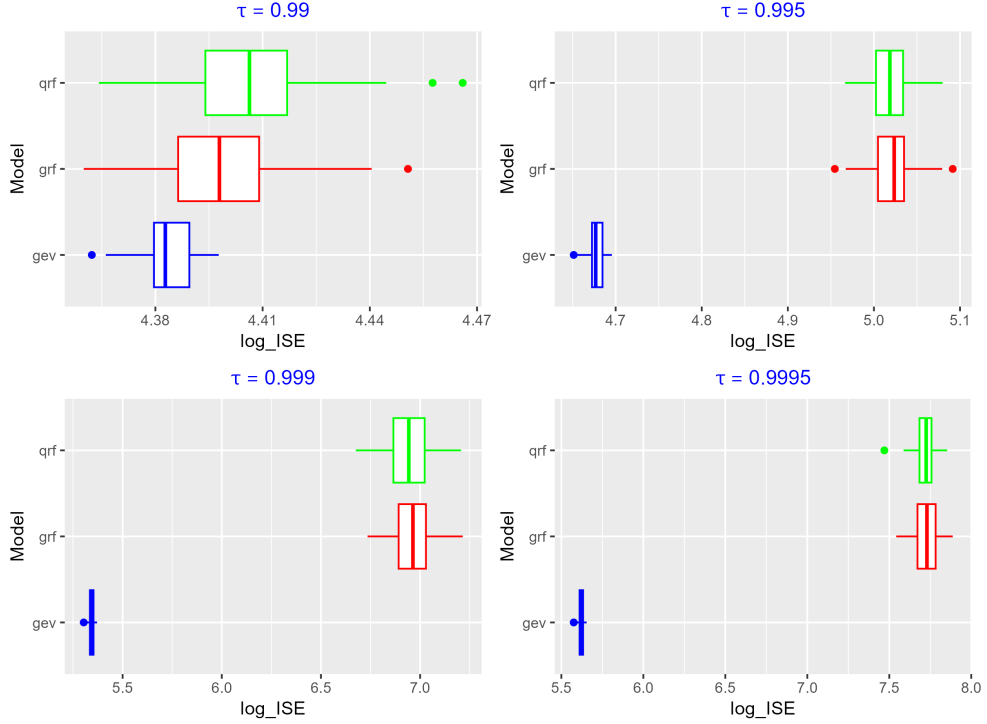
**Figure 2** Boxplot of  $\log(ISE)$  over 100 replication, for  $p = 40$  and  $\tau = (0.99, 0.995, 0.999, 0.9995)$

method’s ability to effectively predict the conditional quantile of the dependent variable, outperforming the GRF and QRF methods even in the presence of noise and when  $Q_\tau(x)$  is complex. This confirms that our method addresses the previously outlined challenges, providing accurate predictions under difficult conditions where the relationship between the conditional quantile function and the covariates is complex and noisy.

#### 4.5 Performance of GEV-erf with Scenario 3

In this scenario, the covariate size is fixed at  $p = 50$ , and the functions  $\gamma(\cdot)$  and  $\nu(\cdot)$  are chosen to make more complex the relationship between the covariates and the conditional quantile of the dependent variable. This scenario highlights the performance of our GEV-erf method in predicting extreme quantiles when the covariate size is large, the quantile function is complex, and in the presence of noise. Figure 3 displays the boxplot of  $\log(ISE)$  for our model as well as for the GRF and QRF models, across

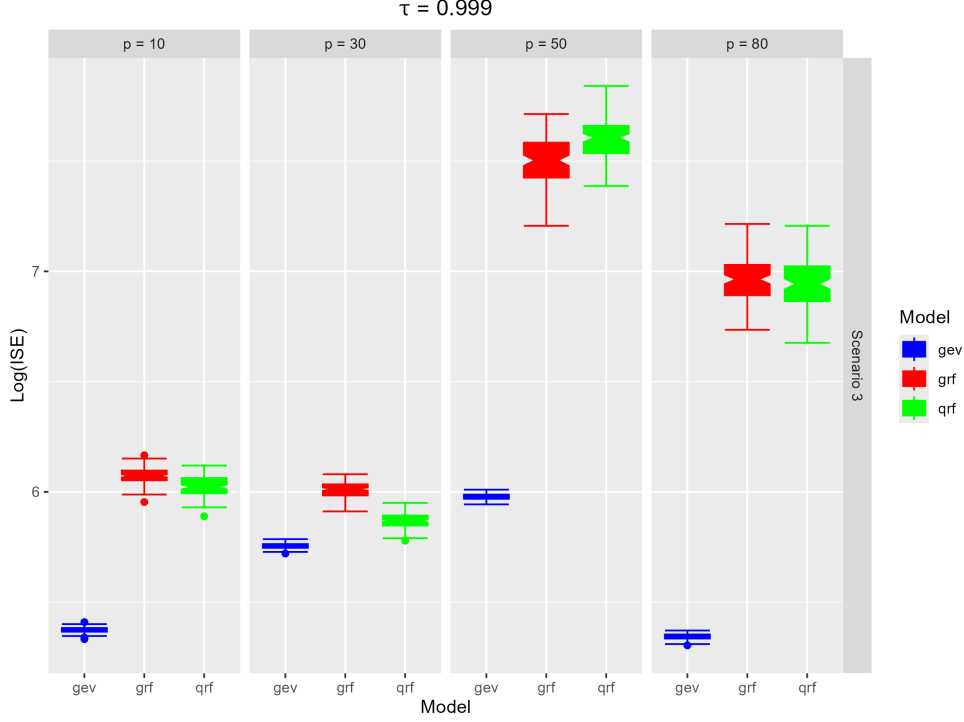
different quantile orders. This figure demonstrates that our GEV-erf method outper-



**Figure 3** Boxplot of  $\log(ISE)$  over 100 replication, for  $p = 50$  and  $\tau = (0.99, 0.995, 0.999, 0.9995)$ .

forms the GRF and QRF methods, indicating its effective capture of the complex structure of extreme quantiles, its adaptation to the high dimensionality of covariates, and its ability to address the challenges outlined in the introduction. In figure 4, we present the boxplot of  $\log(ISE)$  for various methods, considering different covariate sizes and with quantile order set at 0.999. This graph shows that our method is efficient for different values of  $p$ . In terms of performance, our method is superior to GRF and QRF for extreme quantile regression analysis, particularly when the quantile function is complex and the size of the covariates is large. This is particularly evident in scenarios where capturing the subtleties of the quantile function is crucial. The performance of our method remains robust and reliable, highlighting its advantages over other methods.

We evaluate the performance of the models using two additional error metrics: the mean absolute error (MAE) and the median absolute error (MedAE). The results obtained are summarized in Tables 1, 2, and 3, corresponding to the three scenarios of the study. The analysis of these results, conducted for three levels of high quantiles ( $\tau = 0.99, 0.995, 0.9995$ ), demonstrates that the GEV-erf model systematically outperforms the GRF and QRF models. Indeed, GEV-erf exhibits significantly lower errors as well as increased stability, even for extreme quantiles. Conversely, the GRF and QRF



**Figure 4** Boxplot of  $\log(ISE)$  over 100 replication, for  $p \in \{10, 30, 50, 80\}$  and  $\tau = 0.999$ .

models show a strong sensitivity to the highest quantiles ( $\tau = 0.9995$ ), resulting in a marked increase in errors. Moreover, the detailed analysis by scenario confirms this trend: in scenario 1, GEV-erf maintains stable performance, while the errors for the GRF and QRF models increase sharply for extreme quantiles; in scenario 2, a general degradation in performance is observed, but GEV-erf retains relatively higher accuracy; finally, in scenario 3, overall performance improves for all models, with GEV-erf nonetheless confirming its superiority. In summary, the analysis highlights the robustness and precision of the GEV-erf model for estimating high quantiles, while the GRF and QRF models reveal their limitations in dealing with extreme quantiles. These findings further support the conclusions previously obtained.

Scenario 1						
Model	MAE			MedAE		
	$\tau = 0.99$	$\tau = 0.995$	$\tau = 0.9995$	$\tau = 0.99$	$\tau = 0.995$	$\tau = 0.9995$
gev	12.4	14.6	24.9	12.3	14.7	25.5
grf	12.8	16.5	71.8	12.8	16.3	31.7
qrf	12.5	15.9	68.7	12.5	15.1	31.1

**Table 1** Performance of models for different metrics.

Scenario 2						
Model	MAE			MedAE		
	$\tau = 0.99$	$\tau = 0.995$	$\tau = 0.9995$	$\tau = 0.99$	$\tau = 0.995$	$\tau = 0.9995$
<b>gev</b>	17.9	21.0	34.8	18.1	21.3	36.5
<b>grf</b>	18.5	22.2	83.7	18.3	21.8	38.7
<b>qrf</b>	18.8	22.9	88.1	18.5	22.0	37.4

**Table 2** Performance of models for different metrics.

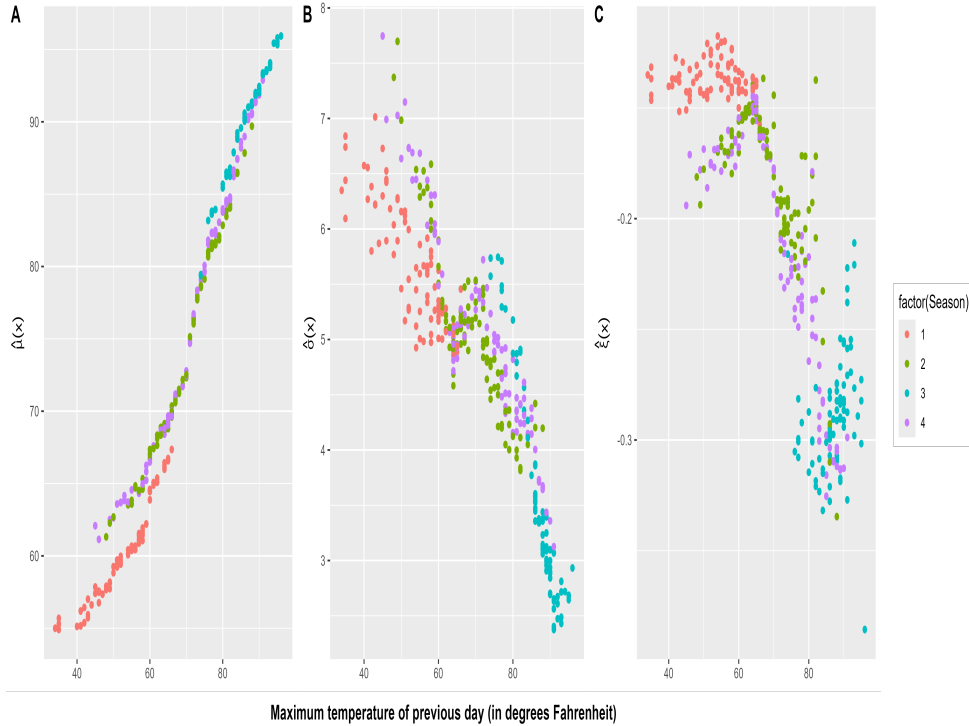
Scenario 3						
Model	MAE			MedAE		
	$\tau = 0.99$	$\tau = 0.995$	$\tau = 0.9995$	$\tau = 0.99$	$\tau = 0.995$	$\tau = 0.9995$
<b>gev</b>	9.50	11.3	19.7	9.08	10.8	19.3
<b>grf</b>	10.1	12.7	36.2	9.85	12.2	33.7
<b>qrf</b>	10.3	12.8	34.1	10.9	11.8	31.7

**Table 3** Performance of models for different metrics.

In the appendix B, we provide additional graphical representations for the different scenarios considered in this work. These graphs validate our conclusions by illustrating the method’s performance for different orders of high quantiles. In so doing, we provide an overview of the efficiency and robustness of our approach in handling extreme quantile regression tasks.

## 5 Applications to real datasets

To evaluate the performance of the extreme quantile regression model GEV-erf in a real-world setting, we use daily meteorological data from the Fort Collins station in Colorado (USA), recorded between January 1, 1900, and December 31, 1999. This dataset, available in [Sielenou and Alain \(2020\)](#), has been used in previous studies such as [Dkengne et al. \(2020\)](#) and [Katz et al. \(2002\)](#). In our analysis, the response variable  $Y$  corresponds to the daily maximum temperature (in degrees Fahrenheit), while the covariates include daily accumulated precipitation (in inches), daily accumulated snowfall, and two transformed variables: the variable *Season*, which takes values 1 (winter), 2 (spring), 3 (summer), and 4 (fall), and the maximum temperature of the previous day. Observations with missing values in any of the variables were excluded, resulting in a total of 35,793 complete observations. To increase the dimensionality of the covariate space and test the robustness of the model in a high-dimensional setting, we add six independent random variables generated from a uniform distribution on  $[-1, 1]$ , bringing the total number of covariates to  $p = 10$ . The analysis focuses on monthly maximum temperatures, obtained by forming monthly blocks from the daily data. For model training, 70% of the monthly maxima are used as the training set, with the remaining 30% reserved for testing.



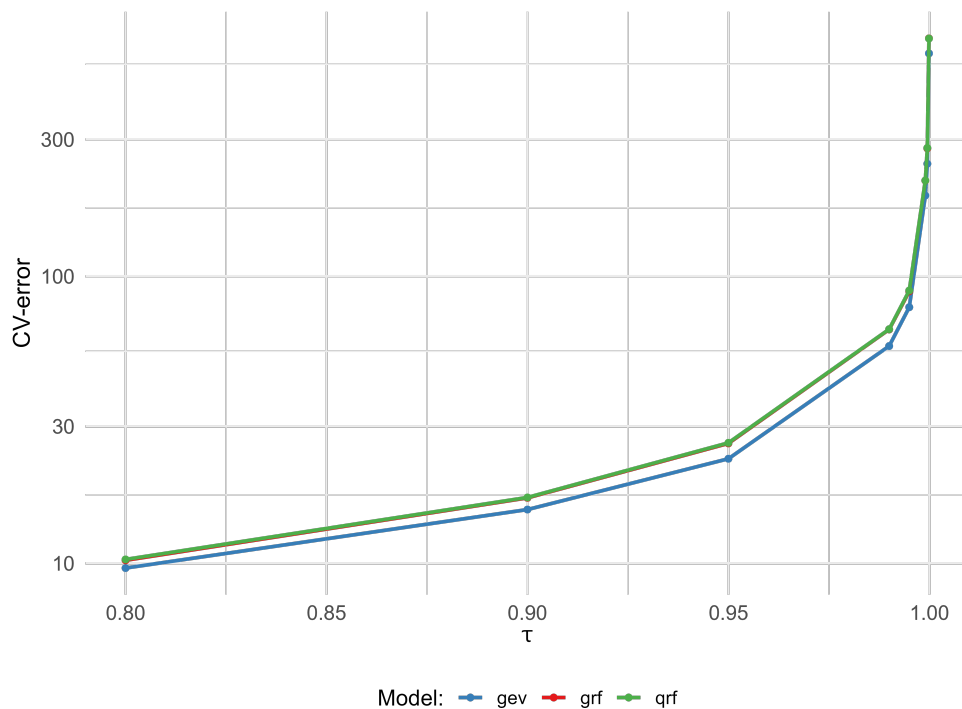
**Figure 5** Variation of the estimated parameters  $\hat{\theta}(x)$  as a function of the previous day’s maximum daily temperature, by season.

The results reveal a strong dependence between the monthly maximum temperature and the previous day’s maximum temperature. Figure 5 illustrates the variation of the estimated GEV distribution parameters  $\hat{\theta}(x)$  with respect to this covariate, with a distinction made across seasons. A nearly linear increasing relationship is observed between the previous day’s maximum temperature and the location parameter  $\hat{\mu}(x)$ , suggesting that warmer preceding days raise the baseline level of expected extreme temperatures. The scale parameter  $\hat{\sigma}(x)$  decreases as the previous day’s temperature increases, indicating reduced variability in the extremes. The shape parameter  $\hat{\xi}(x)$  exhibits a more complex dynamic: for moderate temperatures,  $\hat{\xi}(x)$  remains close to zero, indicating a Gumbel-type tail (moderately heavy). However, for higher temperatures—particularly during summer and autumn— $\hat{\xi}(x)$  becomes significantly negative. This implies a shorter tail, suggesting the presence of an upper bound on extreme temperatures. These findings are consistent with the conclusions of [Dkengne et al. \(2020\)](#), who showed that the unconditional distribution of daily maximum temperature belongs to the Weibull domain of attraction ( $\xi < 0$ ), implying a finite right endpoint. Our conditional modeling using the **GEV-erf** model confirms this property, with estimated values of the shape parameter ranging between  $-0.386$  and  $-0.118$ , as shown in panel C of Figure 5. We conduct a quantitative performance evaluation of the **GEV-erf** method by comparing it with the **GRF** approach of [Athey et al. \(2019\)](#)

and the QRF method of [Meinshausen and Ridgeway \(2006\)](#), using the metric proposed by [Wang and Li \(2013\)](#), defined as follows:

$$R_{n'}(\hat{Q}_x(\tau)) = \frac{\sum_{i=1}^{n'} \mathbb{1}\{Y_i < \hat{Q}_{X_i}(\tau)\} - n'\tau}{\sqrt{n'\tau(1-\tau)}}, \quad (13)$$

where the function  $\hat{Q}_x(\tau)$  denotes the estimated conditional quantile, evaluated on the test sample  $(x_i, y_i)_{i=1, \dots, n'}$ . The predictive performance of the different models is estimated using 5-fold cross-validation. As an evaluation criterion, we use the absolute value of the metric defined in (13). This procedure is repeated 20 times, and the average of the resulting errors is computed to stabilize the evaluation. [Figure 6](#) illustrates the evolution of the average prediction error as a function of the probability level (on a logarithmic scale) for each of the evaluated models. The results clearly show that the **GEV-erf** method outperforms the GRF and QRF approaches, the latter two exhibiting very similar performance regardless of the extreme probability level considered.



**Figure 6** Average prediction error for the different models as a function of the extreme quantile level.

## 6 Conclusion

Extreme quantile regression is a powerful statistical tool that allows for the analysis and prediction of behaviours in the tails of distributions, where rare and extreme events occur. The existing literature on modeling methods for this approach is limited, particularly in the context of using the Block Maxima approach to ensure tail extrapolation. Most available methods use the Peak-Over-Threshold approach (Pasche and Engelke (2024), Velthoen et al. (2023), Farkas et al. (2024), Gnecco et al. (2024)). In this work, we propose a flexible quantile regression method based on the BM approach and the generalized random forest method to address the issues encountered with classical quantile regression methods, primarily the limitation to low-dimensional covariate spaces and the potential complexity of the quantile function. Our proposed method effectively addresses these issues. Using the BM approach, we model the tail of the conditional distribution  $Y|X = x$  for the generalized extreme value distribution with parameters dependent on covariates. These parameters are estimated using a penalized weighted maximum likelihood method, with weights obtained through the generalized random forest method. Simulation studies and application to real dataset show that our method better captures the complex structure of the quantile function and provides good estimates even when the characteristic variables are high-dimensional and in the presence of noise. Our method demonstrates strong performance compared to other quantile regression methods that utilize learning algorithms, such as Meinshausen and Ridgeway (2006) and Athey et al. (2019) method. While this work is more application-oriented, future research should aim to theoretically prove the consistence of our method, as done by Gnecco et al. (2024) for the POT approach. Another perspective is to explore the multidimensional case of  $Y$  (i.e.,  $Y \in \mathcal{Y} \subset \mathbb{R}^q$  with  $q > 1$ ) and examine the spatial aspect of our method.

## Disclosure Statement

No potential conflict of interest was reported by the authors.

## References

- Athey, S., Tibshirani, J., Wager, S.: Generalized random forests. *The Annals of Statistics* **47**(2), 1148–1178 (2019) <https://doi.org/10.1214/18-AOS1709>
- Bücher, A., Lillenthal, J., Kinsvater, P., Fried, R.: Penalized quasi-maximum likelihood estimation for extreme value models with application to flood frequency analysis. *Extremes* **24**(2), 325–348 (2021) <https://doi.org/10.1007/s10687-020-00379-y>
- Breiman, L.: Bagging predictors. *Machine Learning* **24**(2), 123–140 (1996) <https://doi.org/10.1007/BF00058655>
- Breiman, L.: Random forests. *Machine learning* **45**, 5–32 (2001). Publisher: Springer
- Bücher, A., Segers, J.: On the maximum likelihood estimator for the Generalized

- Extreme-Value distribution. *Extremes* **20**(4), 839–872 (2017) <https://doi.org/10.1007/s10687-017-0292-6> . Accessed 2023-10-03
- Cervantes, M.n.N., Benito, S., López-Martín, C.: Assessing the Performance of the Block Maxima Method in Estimating Market Risk. *Research Square Preprint*. ISSN: 2693-5015 (2024). <https://doi.org/10.21203/rs.3.rs-4426771/v1>
- Chernozhukov, V.: Extremal quantile regression. *Annals of Statistics* **33**(2), 806–839 (2005). Publisher: JSTOR
- Coles, S.: *An Introduction to Statistical Modeling of Extreme Values*. Springer Series in Statistics. Springer, London (2001). <https://doi.org/10.1007/978-1-4471-3675-0>
- Dombry, C., Ferreira, A.: Maximum likelihood estimators based on the block maxima method. *Bernoulli* **25**(3), 1690–1723 (2019) <https://doi.org/10.3150/18-BEJ1032> . Accessed 2023-12-10
- Dkengne, P.S., Girard, S., Ahiad, S.: An automatic procedure to select a block size in the continuous generalized extreme value model estimation. *Hal preprint* (2020). Accessed 2025-05-02
- Daouia, A., Gardes, L., Girard, S.: On kernel smoothing for extremal quantile regression **19**(5B) (2013)
- Haan, L., Ferreira, A.: *Extreme Value Theory*. Springer Series in Operations Research and Financial Engineering. Springer, New York, NY (2006). <https://doi.org/10.1007/0-387-34471-3>
- Dombry, C.: Existence and consistency of the maximum likelihood estimators for the extreme value index within the block maxima framework. *Bernoulli* **21**(1), 420–436 (2015) <https://doi.org/10.3150/13-BEJ573> . Accessed 2023-10-11
- El Methni, J., Stupfler, G.: Extreme versions of Wang risk measures and their estimation for heavy-tailed distributions. *Statistica Sinica* **27**(2), 907–930 (2017)
- Ferreira, A., De Haan, L.: On the block maxima method in extreme value theory: PWM estimators. *The Annals of statistics* **43**(1), 276–298 (2015). Publisher: Institute of Mathematical Statistics
- Farkas, S., Heranval, A., Lopez, O., Thomas, M.: Generalized pareto regression trees for extreme event analysis. *Extremes* (2024) <https://doi.org/10.1007/s10687-024-00485-1> . Accessed 2024-05-19
- Fisher, R.A., Tippett, L.H.C.: Limiting forms of the frequency distribution of the largest or smallest member of a sample. *Mathematical Proceedings of the Cambridge Philosophical Society* **24**(2), 180–190 (1928) <https://doi.org/10.1017/S0305004100015681> . Publisher: Cambridge University Press

- Gnedenko, B.: Sur La Distribution Limite Du Terme Maximum D'Une Serie Aleatoire. *Annals of Mathematics* **44**(3), 423–453 (1943) <https://doi.org/10.2307/1968974>
- Gardes, L., Stupfler, G.: Estimating extreme quantiles under random truncation. *TEST* **24**(2), 207–227 (2015) <https://doi.org/10.1007/s11749-014-0403-5> . Accessed 2023-09-26
- Gardes, L., Stupfler, G.: An integrated functional Weissman estimator for conditional extreme quantiles. *REVSTAT-Statistical Journal* **17**(1), 109–144 (2019). Accessed 2023-09-26
- Gnecco, N., Terefe, E.M., Engelke, S.: Extremal Random Forests. *Journal of the American Statistical Association*, 1–24 (2024) <https://doi.org/10.1080/01621459.2023.2300522> . Accessed 2024-01-17
- Gumbel, E.J.: Statistical Forecast of Droughts. *International Association of Scientific Hydrology. Bulletin* **8**(1), 5–23 (1963) <https://doi.org/10.1080/02626666309493293> . Accessed 2023-09-26
- Halton, J.H.: Algorithm 247: Radical-inverse quasi-random point sequence. *Communications of the ACM* **7**(12), 701–702 (1964) <https://doi.org/10.1145/355588.365104> . Accessed 2024-06-11
- Hastie, T.: *The Elements of Statistical Learning*, Second edition edn. Springer Series in Statistics, vol. 2. Springer, New York, NY (2017)
- Hu, G., Franzke, C.L.E.: Evaluation of Daily Precipitation Extremes in Reanalysis and Gridded Observation-Based Data Sets Over Germany. *Geophysical Research Letters* **47**(18), 2020–089624 (2020) <https://doi.org/10.1029/2020GL089624>
- Jenkinson, A.F.: The frequency distribution of the annual maximum (or minimum) values of meteorological elements. *Quarterly Journal of the Royal Meteorological Society* **81**(348), 158–171 (1955) <https://doi.org/10.1002/qj.49708134804>
- Koenker, R., Bassett, G.: Regression Quantiles. *Econometrica* **46**(1), 33 (1978) <https://doi.org/10.2307/1913643>
- Koenker, R., Hallock, K.F.: Quantile regression. *Journal of economic perspectives* **15**(4), 143–156 (2001). Publisher: American Economic Association
- Koenker, R.: Additive models for quantile regression: Model selection and confidence band-aids. *Brazilian Journal of Probability and Statistics* **25**(3), 239–262 (2011) <https://doi.org/10.1214/10-BJPS131> . Publisher: Brazilian Statistical Association. Accessed 2025-05-22
- Katz, R.W., Parlange, M.B., Naveau, P.: Statistics of extremes in hydrology. *Advances in Water Resources* **25**(8-12), 1287–1304 (2002) <https://doi.org/10.1016/>

S0309-1708(02)00056-8

- Laksaci, A., Maref, F.: Estimation non paramétrique de quantiles conditionnels pour des variables fonctionnelles spatialement dépendantes. *Comptes Rendus Mathématique* **347**(17-18), 1075–1080 (2009)
- Meinshausen, N., Ridgeway, G.: Quantile regression forests. *Journal of machine learning research* **7**(6), 983–999 (2006)
- Pasche, O.C., Engelke, S.: Neural networks for extreme quantile regression with an application to forecasting of flood risk. *The Annals of Applied Statistics* **18**(4), 2818–2839 (2024) <https://doi.org/10.1214/24-AOAS1907>
- Sielenou, D., Alain, P.: FCwx.csv **1** (2020) <https://doi.org/10.17632/32w7fw9ysm.1> . Publisher: Mendeley Data. Accessed 2025-05-05
- Smith, R.L.: Maximum likelihood estimation in a class of nonregular cases. *Biometrika* **72**(1), 67–90 (1985) <https://doi.org/10.1093/biomet/72.1.67> . Accessed 2023-12-03
- Takeuchi, I., Le, Q., Sears, T., Smola, A.: Nonparametric quantile estimation. *Journal of Machine Learning Research* **7**(45), 1231–1264 (2006)
- Velthoen, J., Dombry, C., Cai, J.-J., Engelke, S.: Gradient boosting for extreme quantile regression. *Extremes* (2023) <https://doi.org/10.1007/s10687-023-00473-x> . Accessed 2023-09-26
- Wang, H.J., Li, D.: Estimation of Extreme Conditional Quantiles Through Power Transformation. *Journal of the American Statistical Association* **108**(503), 1062–1074 (2013) <https://doi.org/10.1080/01621459.2013.820134> . Accessed 2025-01-22
- Wang, J., You, S., Wu, Y., Zhang, Y., Bin, S.: A Method of Selecting the Block Size of BMM for Estimating Extreme Loads in Engineering Vehicles. *Mathematical Problems in Engineering* **2016**, 1–9 (2016) <https://doi.org/10.1155/2016/6372197> . Accessed 2025-04-25
- Yu, K., Moyeed, R.A.: Bayesian quantile regression. *Statistics & Probability Letters* **54**(4), 437–447 (2001). Publisher: Elsevier
- Özari, i., Eren, z., Erdoğan, E.: A PROPOSAL METHOD TO SELECT THE OPTIMAL BLOCK SIZE: AN APPLICATION ON FINANCIAL MARKETS. *International Journal of Research in Technology and Management* **4**(2), 1–5 (2018)
- Özari, i., Eren, z., Saygin, H.: A new methodology for the block maxima approach in selecting the optimal block size. *Tehnički vjesnik* **26**(5), 1292–1296 (2019). Accessed 2025-04-27

## Appendix A Selection of parameters $\lambda$ and $\text{min.node.size}$

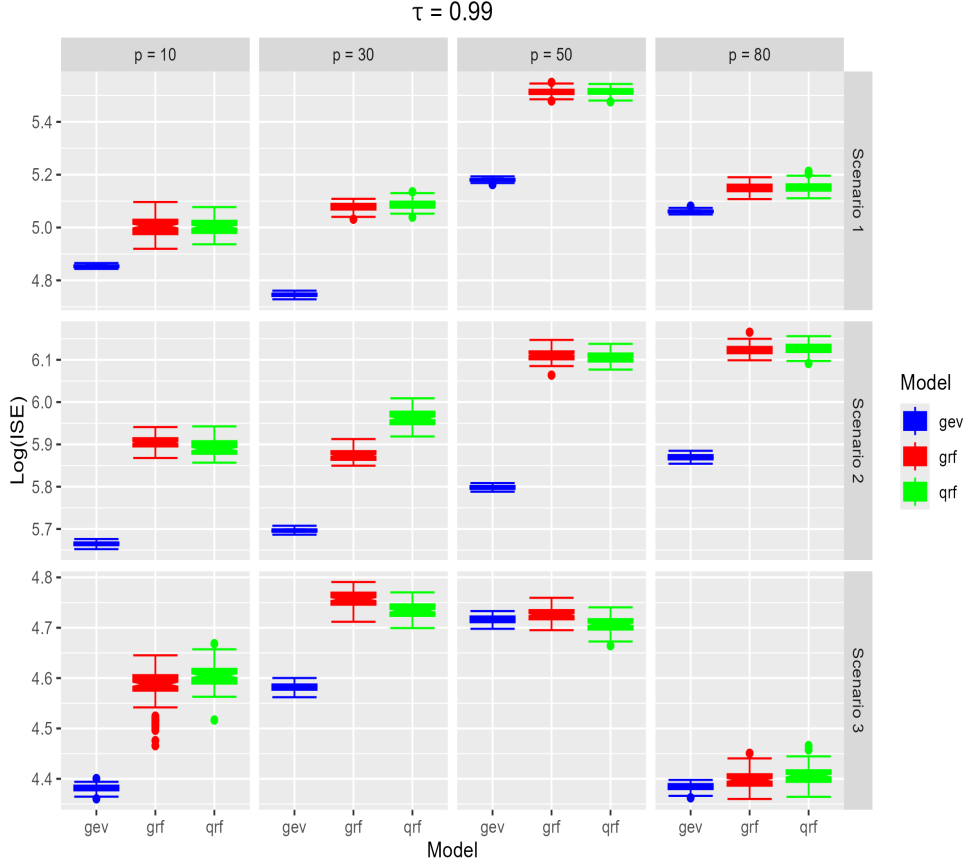
The table (A1) shows the results of the cross-validation used to select the penalty  $\lambda$  and the tuning parameter  $\text{min.node.size}$  for the generalized random forest. To optimize the execution time of the cross-validation, we set the number of trees in the forest to  $\text{num.trees} = 2000$  (the default value in the GRF method). The various  $\lambda$  values tested are  $\{10^{-4}, 10^{-3}, 0.005, 0.01, 0.05, 0.1\}$  and those of  $\text{min.node.size}$  are  $\{5, 10, 20, 50\}$  using 5-fold cross-validation (i.e. we take  $K = 5$  in (12)). We select the pair  $(\lambda, \text{min.node.size})$  that minimizes the cross-validation error defined in (12). The table (A1) shows the cross-validation errors for all combinations of these hyper-parameters in the set under consideration, and for the different scenarios performed. According to this table, the smallest cross-validation error is obtained for the combination  $(\lambda = 0.001; \text{min.node.size} = 10)$  for scenario 1,  $(\lambda = 0.001; \text{min.node.size} = 50)$  for scenario 2 and  $(\lambda = 0.001; \text{min.node.size} = 5)$  for scenario 3.

Grid parameter for CV		CV-Error		
$\lambda$	$\text{min.node.size}$	Scenario 1	Scenario 2	Scenario 3
1e-04	5	11.53729	13.10445	10.42870
1e-03	5	11.53557	13.10410	<b>10.41969</b>
5e-03	5	11.53842	13.10315	10.42660
1e-02	5	11.53971	13.10379	10.42475
5e-02	5	11.53710	13.10341	10.42564
1e-01	5	11.54149	13.10148	10.42543
1e-04	10	11.53785	13.09767	10.42836
1e-03	10	<b>11.53538</b>	13.09863	10.42696
5e-03	10	11.53980	13.09724	10.43051
1e-02	10	11.53842	13.09817	10.42812
5e-02	10	11.53824	13.09793	10.42688
1e-01	10	11.53911	13.09822	10.42852
1e-04	20	11.53810	13.09522	10.43054
1e-03	20	11.53865	13.09498	10.42747
5e-03	20	11.53897	13.09506	10.42786
1e-02	20	11.53886	13.09487	10.43035
5e-02	20	11.53847	13.09440	10.42615
1e-01	20	11.53947	13.09515	10.42804
1e-04	50	11.54066	13.09314	10.42667
1e-03	50	11.53954	<b>13.09209</b>	10.42739
5e-03	50	11.53960	13.09295	10.42588
1e-02	50	11.54043	13.09330	10.42886
5e-02	50	11.54106	13.09272	10.42899
1e-01	50	11.54099	13.09276	10.42816

**Table A1** Adjustment parameters with different combinations of  $\lambda$  and  $\text{min.node.size}$ .

## Appendix B Additional simulation study

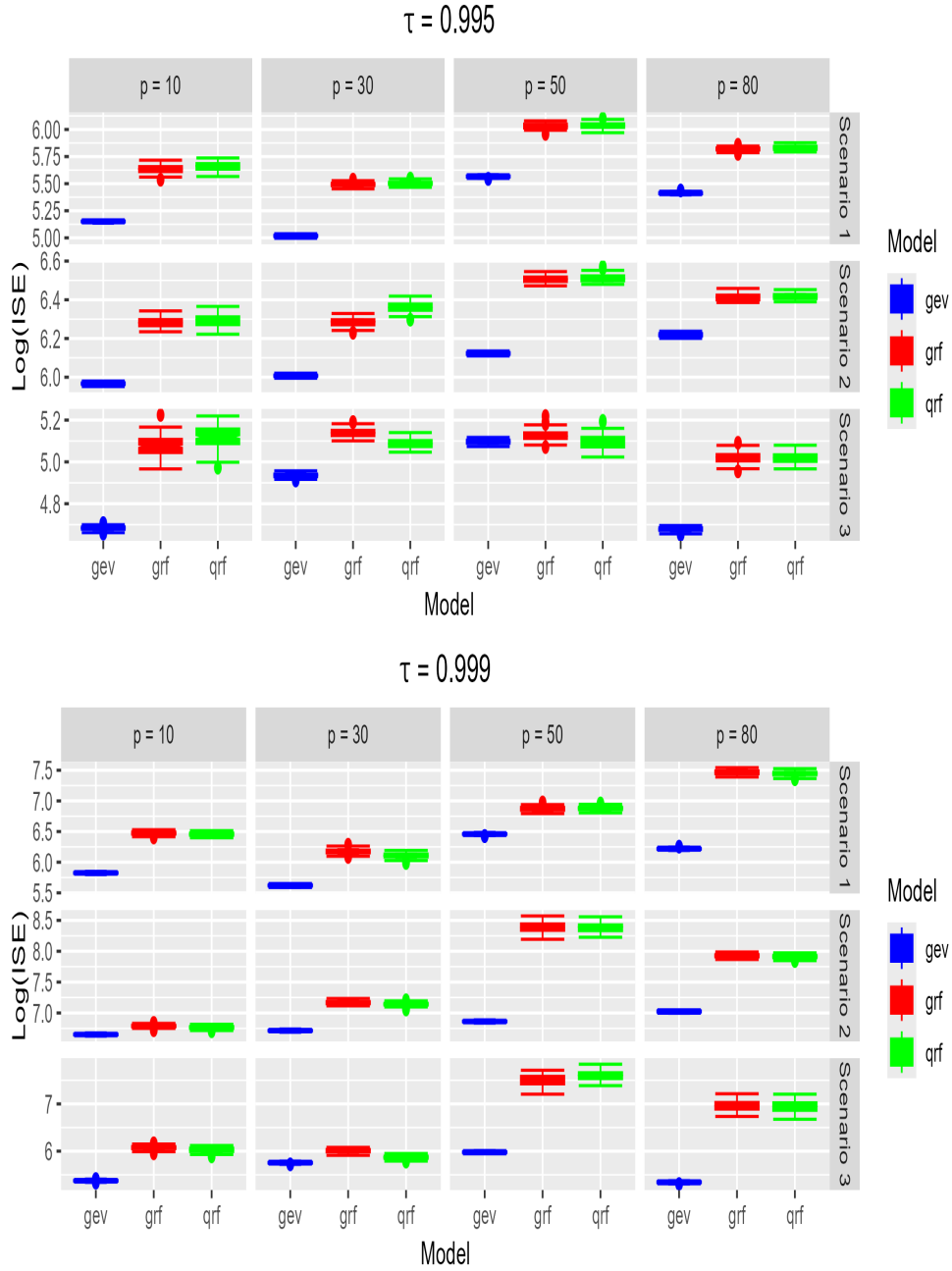
In this section, we perform additional experiments to demonstrate the performance of our method in different scenarios and for various quantile orders. Whatever the quantile order considered, our model maintains high performance across the various scenarios and covariate sizes studied, as shown in Figures B1, B2, and B3.



**Figure B1** Boxplot of  $\log(ISE)$  over 100 replication, for  $p \in \{10, 30, 50, 80\}$ ,  $\tau = 0.99$  and different scenario.

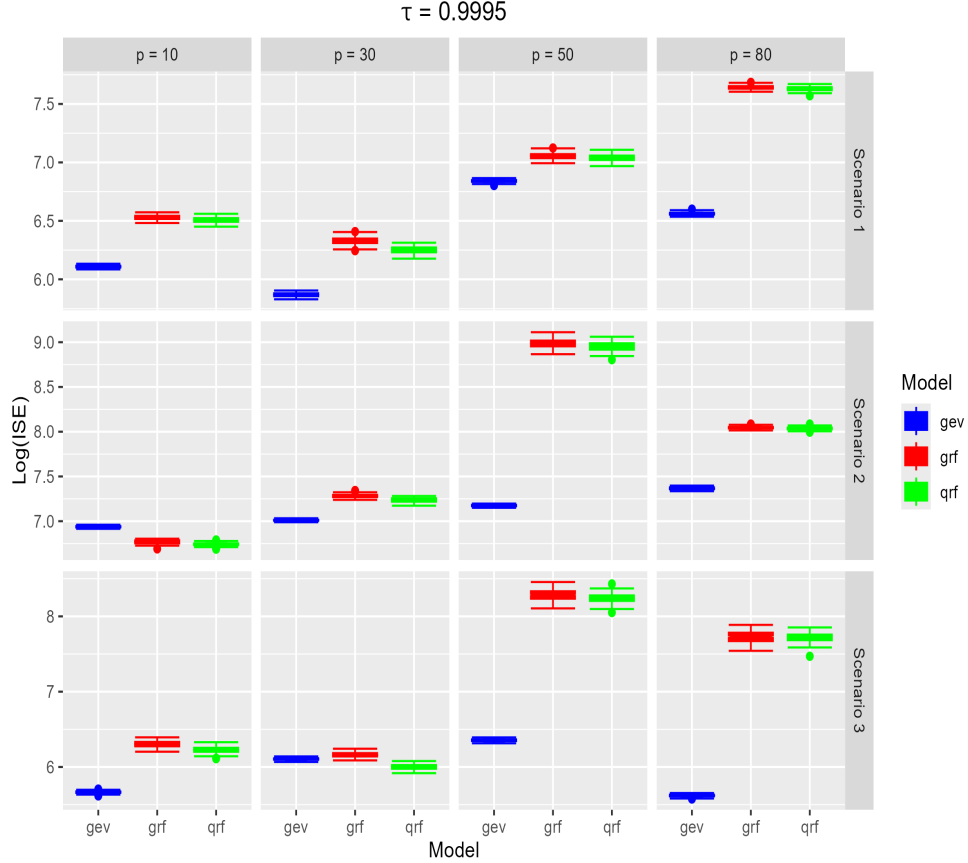
## Appendix C Sensitivity analysis of block size $m$

We study here the sensitivity of the block maxima (BM) method to the block size  $m$ , a crucial parameter for estimating extreme quantiles. An excessively large block size increases the variance of the estimators, while an insufficient block size induces bias. The literature does not provide a universal method for selecting the block size  $m$ , although several recent contributions have attempted to address this issue. For



**Figure B2** Boxplot of  $\log(ISE)$  over 100 replication, for  $p \in \{10, 30, 50, 80\}$ ,  $\tau \in \{0.995, 0.999\}$  and different scenario.

example, [Özari et al. \(2019\)](#); [Özari et al. \(2018\)](#) present a computational approach illustrated by a financial case study. [Wang et al. \(2016\)](#) propose a multi-criteria method



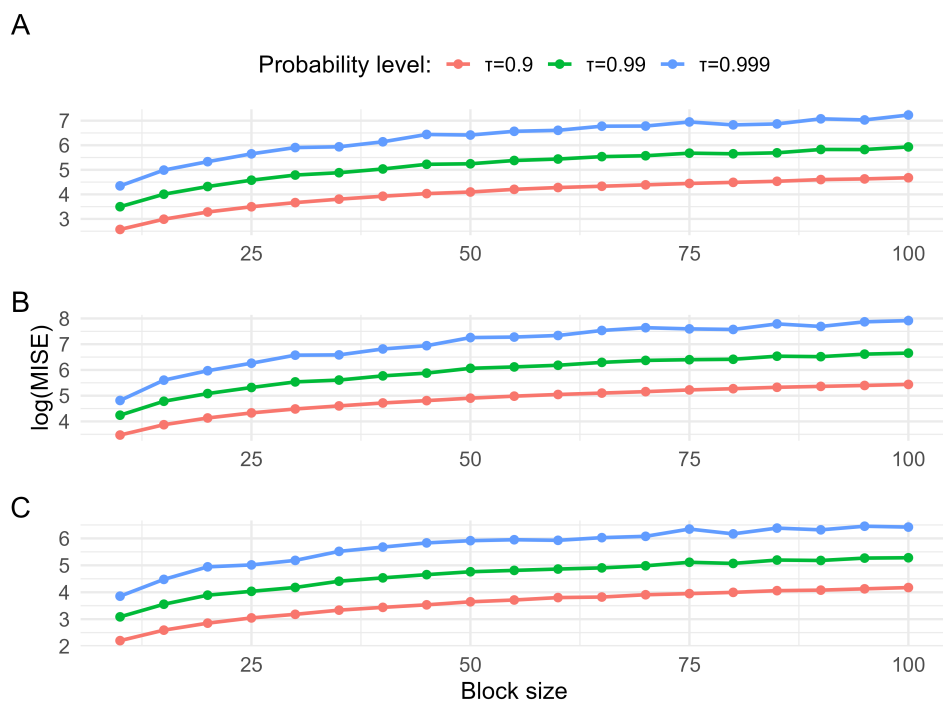
**Figure B3** Boxplot of  $\log(ISE)$  over 100 replication, for  $p \in \{10, 30, 50, 80\}$ ,  $\tau = 0.9995$  and different scenario.

combining graphical analyses and goodness-of-fit tests (Kolmogorov–Smirnov,  $\chi^2$ ). Dkengne et al. (2020) develop an automatic method applied in engineering and meteorology, while Cervantes et al. (2024) assess the fit of the GEV distribution across nine representative block sizes using QQ-plots and statistical tests (Kolmogorov–Smirnov, Anderson–Darling, Cramér–von Mises).

Building on these contributions, we analyze the impact of the block size  $m$  on both the fit of the estimated GEV distribution and the estimation of the conditional quantile. For each scenario, we consider a range of block sizes from 10 to 100 in increments of 5. For each value of  $m$ , the model is trained on a dataset of size  $N$ , and the conditional quantile is then estimated on an independent test sample  $\{(x_i, y_i)\}_{i=1}^{n'}$ , as described in Section 4. The covariate dimension is fixed at  $p = 40$ , the regularization parameter at  $\lambda = 0.001$ , and the parameters *min.node.size* and *num.trees* are retained at their default values from the *grf* package Athey et al. (2019). The GEV-ERF model is fitted to the block maxima samples corresponding to each block size. For each value of  $m$ , two steps are performed using the test sample. First, the MISE (as defined in

Section 4) is computed for various quantile levels ( $\tau \in \{0.9, 0.99, 0.999\}$ ), in order to identify the minimum block size  $m_{\min}$  beyond which the MISE stabilizes. Second, for all  $m \geq m_{\min}$ , we assess the goodness-of-fit of the conditional GEV distribution estimated by the GEV-ERF method using three statistical tests: Kolmogorov–Smirnov (KS), Cramér–von Mises (CVM), and Anderson–Darling (AD). These tests are applied to the probability integral transform of the estimated distribution. More precisely, for each  $i \in \{1, \dots, n'\}$ , we compute  $u_i = G_{\theta(X_i)}(z_i)$ , where  $z_i$  is the block maximum from the test sample, and test whether the  $u_i$  values follow a uniform distribution.

The goodness-of-fit tests are performed using the functions `ad.test` (AD test) and `cvm.test` (CVM test) from the `gofstest` package, and `ks.test` (KS test) from the `stats` package. Lower values of the test statistics indicate a better fit between the model and the data. Figure C4 shows the evolution of the logarithm of the MISE as a function of block size for various quantile levels across the scenarios. The results indicate that while increasing block size tends to raise the MISE of the estimated quantiles, this increase becomes less variable beyond  $m_{\min} = 30$  in Scenarios 1 and 2, and  $m_{\min} = 25$  in Scenario 3. This trend can be explained by the fact that smaller blocks provide more observations for training, thereby improving the performance of the estimation method. Table C2 presents the KS, AD, and CVM test statistics for



**Figure C4** Log(MISE) of Estimated Conditional Quantiles vs. Block Size for scenario 1 (A), scenario 2 (B) and scenario 3 (C)

block sizes  $m \geq m_{\min}$ . The smallest values for the KS and CVM statistics occur around  $m = 45$  in Scenarios 1 (0.061 and 0.172, respectively) and 2. In Scenario 3, the minimum values across all three tests are achieved at  $m = 25$ , indicating a good model fit at that block size. These findings confirm that block size significantly influences the quality of fit of the GEV-ERF model. The results highlight the method’s sensitivity to this parameter and suggest that block sizes in the range of 25 to 50 generally provide a good trade-off between bias and variance, ensuring a satisfactory fit between the estimated GEV distribution and the block maxima. The choice of block size remains a central challenge in extreme value analysis. In practice, it is common to select block sizes based on natural time units—such as a year, season, or month—depending on the temporal resolution of the data.

Block size	Scenario 1			Scenario 2			Scenario 3		
	KS Stat	AD stat	CVM stat	KS stat	AD stat	CVM stat	KS stat	AD stat	CVM stat
25							0,053	1,572	0,234
30	0,072	1,468	0,206	0,046	0,453	0,074	0,080	1,950	0,350
35	0,071	1,290	0,203	0,071	0,950	0,180	0,101	3,845	0,728
40	0,077	1,975	0,296	0,055	0,561	0,088	0,111	3,717	0,704
45	0,061	1,308	0,172	0,036	0,237	0,028	0,119	4,659	0,914
50	0,073	1,571	0,232	0,059	0,468	0,083	0,109	3,523	0,648
55	0,085	1,753	0,245	0,069	0,652	0,103	0,120	3,243	0,573
60	0,090	1,492	0,242	0,055	0,324	0,035	0,103	2,672	0,491
65	0,095	1,686	0,300	0,064	0,451	0,062	0,139	4,437	0,848
70	0,090	1,372	0,223	0,056	0,272	0,031	0,103	1,851	0,307
75	0,102	1,478	0,256	0,047	0,285	0,030	0,097	2,171	0,368
80	0,140	2,296	0,431	0,091	0,980	0,159	0,112	2,876	0,463
85	0,121	2,074	0,385	0,142	2,956	0,442	0,138	3,946	0,735
90	0,102	1,210	0,206	0,058	0,309	0,025	0,128	2,525	0,458
95	0,112	1,647	0,259	0,058	0,281	0,023	0,149	4,444	0,849
100	0,132	2,118	0,388	0,113	1,380	0,176	0,124	1,773	0,306

**Table C2** Statistical values for the various tests as a function of block size (verifying  $m \geq m_{\min}$ ) and scenario.