Nonparametric extrapolation of extreme quantiles: a comparison study

Fabiola Banfi¹ · Greta Cazzaniga¹ · Carlo De Michele¹

Accepted: 21 September 2021 / Published online: 7 October 2021
© The Author(s) 2021

Abstract

The extrapolation of quantiles beyond or below the largest or smallest observation plays an important role in hydrological practice, design of hydraulic structures, water resources management, or risk assessment. Traditionally, extreme quantiles are obtained using parametric methods that require to make an a priori assumption about the distribution that generated the data. This approach has several limitations mainly when applied to the tails of the distribution. Semiparametric or nonparametric methods, on the other hand, allow more flexibility and they may overcome the problems of the parametric approach. Therefore, we present here a comparison between three selected semi/nonparametric methods, namely the methods of Hutson (Stat and Comput, 12(4):331–338, 2002) and Scholz (Nonparametric tail extrapolation. Tech. Rep. ISSTECH-95-014, Boeing Information and Support Services, Seattle, WA, United States of America, 1995) and kernel density estimation. While the first and third methods have already applications in hydrology, Scholz (Nonparametric tail extrapolation. Tech. Rep. ISSTECH-95-014, Boeing Information and Support Services, Seattle, WA, United States of America, 1995) is proposed in this context for the first time. After describing the methods and their applications in hydrology, we compare their performance for different sample lengths and return periods. We use synthetic samples extracted from four distributions whose maxima belong to the Gumbel, Weibull, and Fréchet domain of attraction. Then, the same methods are applied to a real precipitation dataset and compared with a parametric approach. Eventually, a detailed discussion of the results is presented to guide researchers in the choice of the most suitable method. None of the three methods, in fact, outperforms the others; performances, instead, vary greatly with distribution type, return period, and sample size.

Keywords Extreme quantiles · Nonparametric extrapolation · Uncertainty · Order statistics · Kernel density estimation

1 Introduction

The quantile with a level of probability $p \in (0, 1)$ of a random variable X , having a cumulative distribution function F , is defined as $x(p) = \inf\{x : F(x) \geq p\}$. Extreme quantiles (upper and lower) are those beyond the largest or smallest observation, of great interest in hydrological practice (Salvadori et al. 2007), for the design of hydraulic

structures, water resources management, and risk assessment problems (Engeland et al. 2004; Apipattanavis et al. 2010; Volpi 2019). Lower extreme quantiles are used, for example, to monitor droughts, that can be considered the most severe and complex of the different weather-related natural hazards (UNDRR 2019). Drought indicators like the Standardized Precipitation Index (SPI; McKee et al. 1993) define a drought to be mild, moderate, severe, or extreme if return periods are respectively between 2 and 6 years, 6 and 15 years, 15 and 44 years, or higher than 44 years (McKee et al. 1993). Upper extreme quantiles are used, for example, in the design of hydraulic structures. Structures like sewers and dams require design return periods between 2 and 10 years, and between 500 and 10000 years respectively (Chow et al. 1988). Due to the limited length of time series, typically 50–90 years

✉ Fabiola Banfi
fabiola.banfi@polimi.it

Greta Cazzaniga
greta.cazzaniga@polimi.it

¹ Department of Civil and Environmental Engineering,
Politecnico di Milano, Piazza Leonardo da Vinci 32,
Milan 20133, Italy

(Apipattanavis et al. 2010), the quantile of interest is often not observed in the time series (Brazdil and Kundzewicz 2006). When the length of the time series is lower than the return period of interest, extrapolation above (below) the maximum (minimum) observed value becomes necessary.

Traditionally, the extrapolation is obtained adopting a parametric approach. The assumption is made that the variable is described by a given parametric distribution whose set of parameters is estimated from observations. The quantile for a fixed level of probability (or return period) is then obtained inverting the cumulative distribution function (CDF). The limits of this approach, particularly evident when dealing with the tail of the distribution, were highlighted by several authors (e.g., Adamowski 1985; Moon and Lall 1994; Apipattanavis et al. 2010; Lekina et al. 2014). A non appropriate choice of the parametric distribution leads to biases in the estimates. Additionally, data may have a structure difficult to be reproduced with a single parametric function. Extreme floods are often caused by different processes, therefore they require the use of a mixed distribution to be modelled with a parametric approach (Waylen and Woo 1982; Alila and Mtraoui 2002; Carreau et al. 2009; Evin et al. 2011; Barth et al. 2019; De Michele 2019). The choice of the distribution may be driven by different criteria: (1) guidelines (e.g., log-Pearson type III is the standard for annual maximum flood in the United States, see England et al. 2019); (2) the (presumed or real) asymptotic behavior of the variable (e.g., Gamma distribution is commonly adopted to fit precipitation distribution, see Martinez-Villalobos and Neelin 2019); (3) chosen among a pool of distributions (e.g., Sol'áková et al. 2014; Nashwan et al. 2018). In order to compare them, goodness of fit tests are usually applied. Nevertheless, conventional tests look essentially at the fit of the bulk of observations to the model; they do not ensure that extreme quantiles are reproduced adequately (El Adlouni et al. 2008; Brennan et al. 2017); see also the check made on upper quantiles in De Michele and Avanzi (2018). Eventually, simple parametric models may not be able to provide an adequate description of the whole range of data, resulting in a good fit of the body but a non accurate description of the tails.

An answer to these problems may be to resort to semiparametric or nonparametric methods. In both typologies, no distributional assumption is made, but the former still maintains some parameters that require to be selected or estimated. The first category includes kernel estimators (e.g., Silverman 1986), whose basic idea is to obtain a local approximation of the target function using the nearby observations weighted depending on their distance (Lall 1995). Kernel estimators are applied both to fit the probability density function (PDF) or the CDF and the quantile. The earliest applications of this type of semiparametric

methods in hydrology can be found in flood frequency analysis (e.g., Adamowski 1985; Bardsley 1989; Adamowski and Feluch 1990; Lall et al. 1993; Moon et al. 1993; Moon and Lall 1994) due to the inadequacy of classical parametric distributions to model streamflow (Lall and Rajagopalan 2016). Kim and Heo (2002), comparing parametric models and kernel density estimation (KDE), suggested the use of the latter for flood quantile estimation in the interpolation range. Nevertheless, the classical kernel quantile and density estimators have some limitations in extrapolating outside the range of observations (Apipattanavis et al. 2010; Wei et al. 2015). To improve extrapolation performances, Apipattanavis et al. (2010) proposed an estimator based on local polynomial regression and they applied it to flood data. Another possibility to improve tail extrapolation of semiparametric models is to describe the bulk and the tail of the data with different models and to allow a smooth transition between them. Tencaliec et al. (2020) for example proposed a semiparametric version of the Extended Generalized Pareto distributions (EGPDs) to fit the whole range of rainfall data, while MacDonald et al. (2011) proposed a mixture model that combines a kernel density estimator with a tail model in compliance with Extreme Value Theory (EVT). Hutson (2000), instead, combined a parametric model for the tail with a linear interpolation for the central part of the data.

Another type of quantile estimator is the one proposed by Hutson (2002) and already applied in hydrology in the works of Serinaldi (2009, 2011; Serinaldi et al. 2012) and Bezak et al. (2017, 2018, 2019). This method is also at the base of the semiparametric tail-extrapolated quantile estimators proposed by Wei et al. (2015). This latter class of estimators improves traditional quantile estimators (e.g. kernel estimator) in the extrapolation range. The methods proposed by both Hutson (2002) and Wei et al. (2015) are particularly suited for finite sample sizes. A third group of methods is the one based on EVT. In addition to the parametric tail estimation with the method of block maxima or the peak over threshold approach, several semiparametric and nonparametric methods are available in literature (e.g., Hill 1975; Pickands 1975; Dekkers et al. 1989; Beirlant et al. 2005; Lekina et al. 2014). We refer to Beirlant et al. (2004) for a detailed illustration of this class of methods. We cite here also the method proposed by Scholz (1995). In this case EVT is used in order to find a suitable probability plot where the order statistics can be considered to follow a straight line and where quantiles and confidence intervals are extrapolated. Whilst these methods have the advantage of estimating the tail of the distribution looking only at the extreme values, the choice of how many of the k -largest observations should be considered extreme is not straightforward (Beirlant et al. 2004). Besides,

problems may arise in case of small sample sizes since EVT may not hold (Young and Mathew 2014).

Despite the limitations of parametric methods to extrapolate extreme quantiles and the potentiality of semi/nonparametric ones, the latter are not yet widely used in hydrology. Hence, our decision to present a comparison between a selection of them. This work is not intended to be an exhaustive review of semi/nonparametric methods, rather it gives an overview of these tools and their performances under different distributions and sample sizes in order to promote and ease their use in hydrology. The chosen methods, namely Hutson (2002), Scholz (1995), and KDE are illustrated in Sect. 2. Section 3 explains the setup of the simulation study, while in Sects. 4 and 5 results of the simulation study and the case study are respectively presented. Discussions and conclusions are given in Sects. 6 and 7.

2 Methods

2.1 Hutson (2002) method

Hutson (2002) presents a new simple nonparametric quantile function estimator that is essentially an improvement of the linear interpolation estimator, firstly proposed by Parzen (1979). The linear interpolation estimator was developed in order to provide estimations of quantiles falling within the range of observations. Assuming a set of n i.i.d. random variables $\{X_1, \dots, X_n\}$ with support in $\mathbb{R}$ and the order statistics for the sample $\{Y_1 \geq \dots \geq Y_n\}$, the linear interpolation estimator is defined as

$$\hat{x}(p) = (1 - \epsilon) Y_{n-k+2} + \epsilon Y_{n-k+1}, \quad (1)$$

$$\frac{k}{n+1} < p < \frac{k+1}{n+1}, \quad k = 1, \dots, n,$$

where $\epsilon = n'p - \lfloor n'p \rfloor$, $n' = n + 1$ and $\lfloor \cdot \rfloor$ is the floor function. The enhancement proposed in Hutson (2002) consists of allowing the nonparametric quantile extrapolation beyond the range of observed data. The quantile function estimator by Hutson (2002), for a given probability level p , has the following expression:

$$\hat{x}(p) = \begin{cases} Y_n + (Y_{n-1} - Y_n) \log(n'p) & \text{if } 0 < p \leq \frac{1}{n+1}; \\ (1 - \epsilon) Y_{n-k+2} + \epsilon Y_{n-k+1} & \text{if } \frac{k}{n+1} < p < \frac{k+1}{n+1}, \quad k = 1, \dots, n; \\ Y_1 - (Y_1 - Y_2) \log(n'(1-p)) & \text{if } \frac{n}{n+1} \leq p < 1. \end{cases} \quad (2)$$

The equation in the second row represents the linear interpolation estimator by Parzen (1979), i.e. Eq. (1), and it applies to quantiles within the range of probabilities $[\frac{1}{n+1}, \frac{n}{n+1}]$. Expressions in first and third rows represent the

extension of the quantile estimator, proposed by Hutson (2002), for extrapolation of tail extremes outside the range of observed data. The new quantile expression in Eq. (2) is based on the following hypotheses:

1. the image of $x(p)$ is $\mathbb{R}$;
2. $x(p)$ is C^0 in $p = \frac{1}{n+1}$ and in $p = \frac{n}{n+1}$;
3. $x(p)$ is C^1 in $p = \frac{1}{n+1}$ and in $p = \frac{n}{n+1}$.

Finally, there could be cases where the observed variables have support in $\mathbb{R}^+$, this could be the case of rainfall or discharge data. In order to take into account also these conditions, Hutson (2002) implemented a slightly modified version of Eq. (2), here reported:

$$\hat{x}(p) = \begin{cases} \epsilon Y_n & \text{if } 0 < p \leq \frac{1}{n+1}; \\ (1 - \epsilon) Y_{n-k+2} + \epsilon Y_{n-k+1} & \text{if } \frac{k}{n+1} < p < \frac{k+1}{n+1}, \quad k = 1, \dots, n; \\ Y_1 - (Y_1 - Y_2) \log(n'(1-p)) & \text{if } \frac{n}{n+1} \leq p < 1. \end{cases} \quad (3)$$

The same idea, i.e. using a straight line approximation for the bounded tail, can be applied, and the system consequently modified, also in case of upper bounded data. Hutson (2002) carried out simulations with small samples of length n equal to 10 and 25, in order to assess the estimator behaviour. The samples were extracted by Normal, Cauchy, and Exponential distributions. Results showed that, for small sample lengths, the estimator of Hutson (2002) provides better performance with respect to other quantile function estimators, such as the kernel quantile estimator. However, it was proved that the method has difficulties in describing the tail behaviour of heavy-tailed distributions, such as the Cauchy distribution. This method was hence recommended for mid- and light-tailed distributions.

2.2 Scholz (1995) method

The method proposed by Scholz (1995) starts from the knowledge that order statistics can serve as confidence bounds for the population quantiles and it relies on two assumptions regarding the sampled distribution: (1) it is continuous; (2) it belongs to the domain of attraction of one of the three types of asymptotic distributions of extremes. Let $\{X_1, \dots, X_n\}$ be n i.i.d. random variables with distribution $F_X(x)$ and assume that data are realizations of this sample. The sample, ordered in descending order, is then denoted by $\{Y_1 \geq \dots \geq Y_n\}$. For a given p , the ordered observation Y_i can serve as 100% upper confidence bound for the quantile x_p as follows:

$$\alpha = P(Y_i \geq x_p) = \sum_{j=i}^n \binom{n}{j} (1-p)^j (p)^{n-j}. \quad (4)$$

The survival function of Y_i can also be rewritten, using the Beta and incomplete Beta function, as follows:

$$\begin{aligned} \alpha = P(Y_i \geq x_p) &= \frac{\int_0^{1-p} t^{i-1} (1-t)^{n-i} dt}{\int_0^1 t^{i-1} (1-t)^{n-i} dt} \\ &= \frac{\Gamma(n+1)}{\Gamma(i)\Gamma(n-i+1)} \int_0^{1-p} t^{i-1} (1-t)^{n-i} dt. \end{aligned} \quad (5)$$

Since it is not always possible, fixing α and p , to find a choice of i so that Eq. (4) is satisfied closely enough, Scholz (1995) proposed to fix i and α and to compute the associated $p_{i,\alpha,n}$. For $\alpha = 0.5$ a good approximation is given by

$$p_{i,5,n} \approx 1 - \frac{i-1/3}{n+1/3}. \quad (6)$$

The idea is then to find a suitable transformation $g(p)$ so that the relation between $g(p_{i,\alpha,n})$ and Y_i for $i = 1, 2, \dots, k$ can be assumed linear and to extrapolate the tail of the distribution in that plane. The value of k gives the depth of the sample used to extrapolate the extreme quantiles. We will focus in the following only on the 50% confidence bound, since it can be used as a median unbiased estimate of x_p . Making use of EVT and recalling that $F_X(x)$ is in the domain of attraction of the extreme value distribution, for $n \rightarrow \infty$ we have that

$$x(p) \approx \delta \cdot \frac{(-n \ln(p))^{-\gamma} - 1}{\gamma} + \lambda, \quad (7)$$

where $\lambda \in \mathbb{R}$, $\delta > 0$, and $\gamma \in \mathbb{R}$ are the location, scale, and shape parameter. The shape parameter is also called extreme value index (EVI). An appropriate transform may therefore be

$$g(p) = \frac{(-n \ln(p))^{-\gamma} - 1}{\gamma}. \quad (8)$$

The value of γ was obtained with the estimator proposed by Dekkers et al. (1989):

$$\hat{\gamma}_k = M_{1,k} + 1 - 0.5 \left(1 - \frac{M_{1,k}^2}{M_{2,k}} \right)^{-1}, \quad (9)$$

where $M_{1,k} = \frac{1}{k-1} \sum_{i=1}^{k-1} \log(\tilde{Y}_i / \tilde{Y}_k)$, $M_{2,k} = \frac{1}{k-1} \sum_{i=1}^{k-1} \log^2(\tilde{Y}_i / \tilde{Y}_k)$, $\tilde{Y}_i = Y_i - \text{median}(X_1, \dots, X_n)$ and k is the number of order statistics used in the estimation. The choice of Scholz (1995) to subtract the median from Y_i was motivated by two reasons: (1) logarithm argument must be positive; (2) $\hat{\gamma}_k$ becomes location and scale invariant. Once γ is estimated, the

parameters of the linear relation between Y_i and $g_{\hat{\gamma}}(p_i)$ with $i = 1, \dots, k$ are estimated using weighted least squares. The weights are obtained from the covariance matrix of $\mathbf{Y}$:

$$\text{cov}(Y_i, Y_j) \approx \delta^2 \cdot i^{-\gamma-1} j^{-\gamma}. \quad (10)$$

In this way the different variances of Y_i and their correlation are taken into account. We refer to Scholz (1995) for the derivation of Eq. (10) and the solution of least squares. Once the linear relation is known, it is possible, for a given value of p , to estimate the corresponding $x(p)$.

A last comment regards the number of extreme statistics to be included in the fit. Large k will produce higher bias while smaller k will produce higher variance (Beirlant et al. 2004). In this study we did not use the method proposed by Scholz (1995) for the choice of k because it was not suitable for small n . We therefore proceeded in the following way: (1) the value of γ was estimated for each value of k between two limiting values K_1 and K_2 ; (2) for each value of $\hat{\gamma}$ we estimated the parameters of the linear relation, that we will call b_1 and b_2 , and the associated coefficient of determination, R^2 ; (3) we selected the final b_1 and b_2 averaging all the values for which (a) $R^2 > 0$; (b) $R^2 > 0.7$. For K_1 and K_2 we adopted the values proposed by Scholz (1995): $K_1 = \max(6, \lfloor 1.3\sqrt{n} \rfloor)$ and $K_2 = 2 \lfloor \log_{10}(n)\sqrt{n} \rfloor$.

2.3 Kernel density estimation method

The KDE refers to a statistical method providing semi-parametric estimations of PDF, for which a fixed mathematical form is not defined. It can be thought as an alternative method to histogram, providing smoother results (Wilks 2011). KDE makes use of kernel functions, $k(-)$, in a number equal to the one of observed data. Each kernel function has its center in the data value. KDE is therefore calculated by adding the heights of all kernel functions through a linear superposition. Given n i.i.d. random variables $\{X_1, \dots, X_n\}$, the estimated PDF is given by

$$\hat{f}_X(x) = \frac{1}{nh} \sum_{i=1}^n k\left(\frac{x - X_i}{h}\right), \quad (11)$$

where h is called bandwidth or smoothing constant. The kernel function could be any function which has the following three properties:

1. $\int k(x) dx = 1$;
2. $\int x k(x) dx = 0$;
3. $\int x^2 k(x) dx = a < \infty$.

It follows that any symmetric PDF, with finite variance, can be used as kernel. Nevertheless, it is not a necessary condition that the kernel function is a PDF. Among the

most popular and broadly used kernel functions we can mention Gaussian, Triangular, Cosine, Biweight, and Epanechnikov kernels. Properties of KDE are therefore strictly dependent on the choice of two parameters: (1) the shape of the kernel function and (2) its bandwidth h . Many studies were developed in the past to define the optimal type of kernel (e.g., Epanechnikov 1969; Wand and Jones 1994) and the optimal bandwidth (e.g., Silverman 1986; Sheather and Jones 1991; Cheng and Sun 2006). However, it was proved that the selection of kernel has a rather limited effect on the estimator efficiency (Wand and Jones 1994), that is why the choice is typically based on the simplicity of implementation. Conversely, the bandwidth selection is a more critical step (Sheather 2004; Charpentier and Flachaire 2015), determining the amount of smoothness of the estimation: larger bandwidths result in smoother estimators and smaller bandwidths produce rougher estimates.

Once the PDF is estimated, the aim is to evaluate the quantile. From Eq. (11) can be derived the estimate of the distribution function (Moon et al. 1993) as

$$\hat{F}_X(x) = \frac{1}{n} \sum_{i=1}^n k^* \left(\frac{x - X_i}{h} \right), \quad k^*(t) = \int_{-\infty}^t k(u) du. \quad (12)$$

Finally, an estimator of the quantile, $\hat{x}(p)$, can be calculated as the inverse of the kernel distribution function estimator, $\hat{F}_X^{-1}(x)$.

However, some problems may arise when the quantile to be estimated requires an extrapolation. We indeed saw that the kernel smoothing technique relies on the arbitrary choices of h and $k(-)$. This means that, in case of quantile estimations far beyond the range of observed data, only few observations in the interval h will influence the estimation (Adamowski 1989). Moreover, the selected kernel shape may not reflect the upper tail features (Adamowski and Feluch 1990). To counter these limitations, many works in the past tried to provide an enhancement to the classic KDE. Adamowski (1989) proposed a variable kernel estimator (VKE), where h varies in inverse proportion to the local density of observations. Adamowski and Feluch (1990) recommended the use of generalized extreme value distribution (GEV) as kernel function in flood frequency analysis. In the present work we implemented the newly proposed robust kernel function, $k_\omega(t)$, presented by Wang et al. (2020), which is a mixture of a light-tailed kernel, $k_{light}(x)$, and a thick-tailed kernel, $k_{heavy}(x)$:

$$k_\omega(x) = \omega \cdot k_{heavy}(x) + (1 - \omega) \cdot k_{light}(x), \quad \omega \in [0, 1], \quad (13)$$

where the tuning parameter ω governs the relative weight to be assigned to each component. We chose the standard Normal density function as $k_{light}(t)$, as recommended by

authors in Wang et al. (2020). As k_{heavy} we implemented the Cauchy density function since Adamowski (1985) and Lall et al. (1993) selected it as heavy-tailed kernel for flood frequency analysis. The tuning parameter is thereby optimized, together with the bandwidth, through the likelihood cross validation (Duin 1976), formulated for k_ω as

$$\max_{h > 0, \omega \in [0, 1]} \frac{1}{n} \sum_{i=1}^n \log \left[\frac{1}{(n-1)h} \sum_{j=1, j \neq i}^n k_\omega \left(\frac{X_i - X_j}{h} \right) \right]. \quad (14)$$

2.4 Hydrological applications

Hutson (2002) was developed with the aim of improving bootstrap resampling theories in small samples but it can definitely yield benefits in issues concerning extrapolation of hydrological extremes and the associated confidence intervals. However only few works in hydrological literature proposed an application of the method. In Serinaldi (2009) the estimator was applied to compute confidence intervals for extreme quantiles from datasets of annual peak discharge. Later, Serinaldi (2011) implemented it to define semiparametric flow duration curves, annual flow duration curves, and their confidence intervals. Most recent works (e.g., Serinaldi et al. 2012; Bezak et al. 2017, 2018, 2019) took advantage of the method for copula-based models in hydrological modelling.

A much wider dissemination has been experienced regarding kernel estimators in hydrological fields. They were firstly introduced in hydrology with the aim of finding robust techniques capable of estimating relative frequencies of rare floods or low flows, which lie beyond the range of observed data. Adamowski (1985) proposed the kernel function in hydrological issues to estimate the PDF of annual floods. Later, Bardsley (1989) implemented a kernel estimator to estimate the cumulative distribution function with bandwidths varying with the ranked floods. Adamowski and Feluch (1990) showed that including historical information into kernel flood–frequency analysis improves extrapolation. Lall et al. (1993) focused on bandwidths and kernel functions selection and their impact on flood frequency analysis performance while Moon et al. (1993) investigated estimators on upper tail quantiles. Again, in Moon and Lall (1994), authors implemented the kernel quantile estimator, using boundary kernels in case of quantiles extrapolation outside the range of recorded flows. More recently, the kernel smoothing was applied in other hydrological contexts such as interpolation of missing rainfall data (Lee and Kang 2015) or the estimation of return period of droughts (Kim et al. 2003).

At our knowledge, Scholz (1995) has not yet been used in hydrological applications.

3 Setup of simulation study

We investigated and compared behaviour and performances of the three semi/nonparametric extrapolation techniques, illustrated above, within a hydrological framework. To this aim we carried out an extensive simulation analysis trying to cover a variety of cases which could be encountered in hydrological contexts, sampling four well-known distributions with different attraction domains of maxima and using several sample lengths. The four distributions tested are the Gamma, the left-truncated Cauchy, the Uniform, and the Generalized Pareto. The Gamma distribution has a smoothly varying form and it is commonly adopted for description of skewed hydrological variables, such as precipitation depth of wet days (Chow et al. 1988). For example, it is usually used to fit monthly aggregated precipitation data in the computation of the SPI. The distribution of its maxima is shown to converge asymptotically to the Extreme Value Type 1 (EV1) or Gumbel distribution (Salvadori et al. 2007), that is light-tailed. We recall that the Gumbel distribution is characterized by a value of the EVI equal to 0. The left-truncated Cauchy is a right heavy-tailed distribution. In fact, its maxima follow the behaviour of the Fréchet distribution, the Extreme Value Type 2 (EV2) for which $\text{EVI} > 0$ (Salvadori et al. 2007). In hydrology, the left-truncated Cauchy is mainly exploited in flood frequency analysis, to fit the variables that describe maximum annual flood events, namely flood volume, flood duration, and peak flow (e.g., Ghorbani et al. 2011; El Fels et al. 2018; Nashwan et al. 2018). The truncation of the Cauchy distribution results in a function with positive support and this makes it adequate to study hydrological variables, the majority of which can not assume negative values. The Uniform distribution can be considered the simplest (and let's say uninformative) distribution, which is why it is employed in numerous applications, also in reason of the *Probability Integral Transform*, which allows to transform any variable in a Uniform variable in the $[0,1]$ interval. This makes the Uniform distribution fundamental in the copula-based multivariate statistical modelling, since the marginal probability distributions of copula are uniform in $[0,1]$. Besides, Uniform distribution is used for the construction of the unit hydrograph in urban hydrology and can describe river morphology in case of dynamical equilibrium of the river (Singh 1998). Since it is upper and lower bounded, the Uniform distribution belongs to the domain of attraction of Weibull or Extreme Value Type 3 (EV3), with $\text{EVI} < 0$ (Salvadori et al. 2007). Finally, the Generalized Pareto distribution is used to study maxima with the peak over threshold (POT) method; asymptotically, exceedances over a threshold, u , follow the Generalized Pareto distribution

(Salvadori et al. 2007). The distribution of its maxima depends on the value assumed by the shape parameter γ ; for $\gamma = 0$, maxima follow a Gumbel distribution, for $\gamma > 0$ a Fréchet distribution, and for $\gamma < 0$ a Weibull distribution. In the study we simulated samples from a Generalized Pareto with $\gamma > 0$ and with EVI lower than the one of the left-truncated Cauchy. This choice was driven by the consideration that the extrapolation from heavy-tailed distributions is more problematic, therefore having multiple information regarding this typology may provide a better understanding.

In many of the hydrological applications that require the extrapolation of extreme quantiles, the variables of interest are annual maxima or peaks over a threshold. These are traditionally studied using Generalized Extreme Value (GEV) or Generalized Pareto distribution. Nevertheless, we decided not to simulate from the GEV (or also from a Generalized Pareto with γ lower or equal to zero) for the following reasons: (1) EVT holds only asymptotically and this may be a problem when dealing with some variables like maximum annual daily precipitation, since the actual sample size of the blocks (the number of wet days in each year) may be also significantly lower than 365 in arid regions. In this case the use of GEV may be inappropriate (De Michele and Avanzi 2018); (2) in some hydrological applications, like the computation of drought indices, the variables of interest are not maxima; (3) we covered with the distributions chosen the three types of asymptotic behaviors of maxima that can therefore provide guidelines also for other distributions with similar tails.

Table 1 shows a summary of the four selected parametric distributions, with the analytical formula of the PDF and the related maxima attraction domain. In addition, we reported the values of distributions' parameters chosen for the analysis and the theoretical values of EVI associated to the parent distributions. In fact, these last values can be derived analytically as reported in Beirlant et al. (2004) and Salvadori et al. (2007). The parameters of Gamma distribution were chosen after fitting a series of monthly precipitation, the parameters of the left-truncated Cauchy and Generalized Pareto distribution were taken equal to the values estimated by Ghorbani et al. (2011) fitting a series of maximum annual discharge.

From each distribution, 500 samples were generated for each sample size n , with $n = 25, 50, 75, 100, 200, 500$. The sample lengths were chosen keeping in mind the distributions selected and the typical hydrological applications in which quantile extrapolation is required. In drought analysis, for example, from a 50 years daily precipitation series, it is possible to extract a sample of aggregated monthly precipitation with a length of 50 or 600 depending if the data are grouped or not month by month. Series of duration, volume, or peak discharge of maximum flood

Table 1 Description of the four selected parametric distributions and parameters' values chosen for the analysis

Distribution	PDF	Domain	Parameters value	Attraction domain of maxima	Theoretical EVI
Gamma	$f(x) = \frac{1}{\Gamma(k)\Theta^k} x^{k-1} \exp(-\frac{x}{\Theta})$	$x > 0$	$\kappa = 3$ $\Theta = 1$	Gumbel (light tailed)	0
Left-Truncated Cauchy	$f(x) = \frac{1}{\pi s \left[1 + \left(\frac{x-x_0}{s} \right)^2 \right] \left[\frac{1}{2} + \frac{1}{\pi} \tan^{-1} \left(\frac{x_0}{s} \right) \right]}$	$x > 0$	$x_0 = 31.46$ $s = 16.08$	Fréchet (heavy tailed)	1
Uniform	$f(x) = \frac{1}{b-a}$	$x \in [a, b]$	$a = 0$ $b = 1$	Weibull (upper bounded)	-1
Generalized Pareto	$f(x) = \frac{1}{\sigma} \left(1 + \gamma \frac{x-\mu}{\sigma} \right)^{-\frac{1}{\gamma}-1}$	$x > \mu$	$\mu = 0.20$ $\sigma = 5.61$ $\gamma = 0.15$	Fréchet (heavy tailed)	γ

events have a length equal to the number of years in the series, typically 50–90 years, or the number of years $\times$ the number of exceedances, depending if block maxima or POT method is considered. Then, we selected five levels of probability $p = 0.98, 0.99, 0.995, 0.998, 0.999$, or equivalently in terms of the return period $T = 50, 100, 200, 500, 1000$ years (being $p = 1 - \frac{1}{T}$), which are the values commonly used for hydraulic design and risk assessment problems. We finally selected the relative error performance criterion (Scholz and Tjoelker 1995) to evaluate the behaviour of semi/nonparametric extrapolation techniques when applied to the sampled data. It is a location and scale invariant criterion as it does not depend on distribution center's location and its spread. The criterion formulation is

$$\eta(p) = \frac{\hat{x}(p) - x(p)}{x(p) - m(F)}, \quad (15)$$

where $\hat{x}(p)$ is the quantile estimated through one of the three semi/nonparametric techniques and associated to a probability level p , $x(p)$ is the corresponding theoretical value calculated as the inverse of the theoretical cumulative distribution function $F_X(x)$ and $m(F)$ is the median of the sampled distribution.

In the present work we only focused on the extrapolation of maxima, rather than minima. In fact, in reason of the *Symmetry Principle*, the following transformation allows to extend analysis from maxima to minima (Salvadori et al. 2007):

$$\min\{X_1, \dots, X_n\} = -\max\{-X_1, \dots, -X_n\} \quad (16)$$

For this, analysis and discussion that will be reported for maxima could be easily extended to minima.

4 Results of simulation study

In this section we provide an overview of the extrapolation performances that can be achieved through the use of the three selected techniques, with respect to the different cases tested. Figure 1 compares medians and interquartile ranges (IQR) of relative errors, from 500 simulations, for each sample length, and subdivided with respect to the type of distribution and return period of extrapolated quantiles. The acronyms HTS, SC0, and SC7 respectively refer to Hutson (2002), and Scholz (1995) considering results for which the linear interpolation has $R^2 > 0$ and $R^2 > 0.7$. A median error equal to 0 means that half of the simulations provide estimates above the theoretical quantile value and half below it. As median relative error approaches 0 and as interquartile range gets smaller, the performance gets better.

Results show a strong dependence on the type of distribution. We observe mostly negative median errors regarding extrapolated quantiles from the Gamma, the left-truncated Cauchy, and the Generalized Pareto distributions. In contrast, positive median errors affect, for the major part, the estimates from Uniform distribution. The highest median errors, in absolute terms, are observed for the left-truncated Cauchy distribution, reaching almost 100%. Conversely, the Uniform distribution provides the smallest median errors, not higher than 5%, two orders of magnitude smaller than left-truncated Cauchy distribution. Meanwhile, absolute median errors related to Gamma and Generalized Pareto distribution do not exceed 26%. The statistical dispersion of relative error, which is described by the interquartile range, is strictly dependent on the distribution too. Again, the widest dispersion is observed with the left-truncated Cauchy distribution and the narrowest with the Uniform.

Fig. 1 Medians (points) and interquartile ranges (vertical bars) of relative errors, η , of extrapolated extreme quantiles for different sample lengths

In Fig. 2 the performances of quantile extrapolations, aggregated by method, are compared from a different perspective; thus allowing a simpler comparison among different return periods. Here, box plots of relative errors are reported only for synthetic samples with length equal to 50 (see Figs. 6, 7, 8, 9, 10 for results of the remaining sample lengths). Box plots related to Gamma distribution show worse performances for higher return periods; in particular, both the variability and the absolute value of the median error increases for all the methods. A different behaviour is observed in the case of left-truncated Cauchy distribution. For Hutson (2002) and KDE methods, dispersion decreases with the return period but bias increases. Estimates from Scholz (1995) get worse, with regard to both median error and dispersion, as the return period gets larger. Concerning Uniform distribution, in the case of Hutson (2002) and KDE methods, we observe better performances (lower absolute median biases and smaller

IQRs) for higher return periods. On the other hand, for Scholz (1995) higher return periods correspond to worse performances. Finally, for the Generalized Pareto distribution median errors worsen with increasing return period with Hutson (2002) while remaining similar with KDE and Scholz (1995). Interquartile range increases with return period for Scholz (1995) while it does not show appreciable differences for the other methods.

In addition, we can observe from Figs. 1 and 2 different extrapolation performances with respect to the implemented technique. It is not possible to define a preferred one, always providing the best results. Each of them has advantages or drawbacks according to the performed analysis, as shown below.

Fig. 2 Box plots of relative errors, η , of extreme quantiles extrapolated from samples with a length equal to 50

4.1 Hutson (2002) method

The method proposed by Hutson (2002) leads to performances that worsen moving from lighter to heavier tailed distributions. In Fig. 1 we observe the smallest median relative errors for estimates from Uniformly distributed samples, that increase, in absolute values, moving to Gamma and then to Generalized Pareto. The worst are finally gained with left-truncated Cauchy distribution. It can be noticed that, for the first three distributions, extrapolations get better with the increase of sample length. We indeed observe that median relative errors get closer to zero and the variability decreases, with a few exceptions for small return periods. Conversely, in the case of left-truncated Cauchy distribution, the method provides large biases in the quantiles extrapolation, with absolute median errors almost up to 100%. Besides, the interquartile range notably grows, although the median relative error decreases with the increase of sample length.

4.2 Scholz (1995) method

Extrapolation estimates provided by Scholz (1995) rely on the evaluation of two parameters, EVI and k . Figure 3 shows the variability, by the use of box plots, of EVI estimates resulted from 500 simulations, for each distribution and sample length. Box plots exhibit a common tendency of EVI median values to approach the theoretical values (dotted lines) and of interquartile ranges to shrink with the increase of the sample size. This behaviour reflects on extrapolated estimates. In Fig. 1 we indeed observe a general improvement of performance with the increase of sample length. It should be noted that considering only EVI values which provide $R^2 > 0.7$ does not enhance the

estimates, actually it often worsens them. We found synthetic samples for which none of the sample's depths, k , in the range $[K_1, K_2]$, provided a value of R^2 greater than 0.7 in the linear regression. Whenever such situation was met, we replaced the sample with a new one. For Gamma, Uniform and Generalized Pareto distribution, we found out that this condition accounted for less than 1% of the 500 simulations and only for samples with a length equal to 25 and 200. We experienced much more problems when sampling from the left-truncated Cauchy distribution. We thrown out from more than 2% (for samples with sample length equal to 25) to more than 30% (for samples with sample length equal to 500) of the 500 simulations. In the estimate of EVI, it is important to balance between the choice of the threshold to be imposed to R^2 and the number of k contributing to the estimate. Choosing a high R^2 is not always the best choice.

Compared with Hutson (2002) and the KDE, this method provides the best performances, in terms of median relative error, in the case of heavy-tailed Generalized Pareto and left-truncated Cauchy distribution and when sample lengths are equal or higher than 50. The method also gives an excellent performance in the case of Gamma distribution for sample lengths larger than 75. With regard to Uniform distribution, Scholz (1995) provides the worst performance, even though median errors are still small, never higher than 3%.

Applying Scholz (1995), we noticed that least squares could not be solved for some synthetic samples and for some values of k . This was associated with very negative estimates of the EVI that resulted in very high values of the product $i^{-\gamma-1}j^{-\gamma}$ in Eq. (10). In order to apply the method we therefore disregarded all the values of k for which least squares could not be solved.

Fig. 3 Box plots of estimated EVI for different sample lengths

4.3 Kernel density estimation method

Extrapolated quantiles from KDE technique with a mixture kernel depend on optimization of two parameters, ω and h . Box plots of optimized ω , for the 500 simulations, are reported in Fig. 4. Here, high values of ω correspond to heavier tails while lower values to lighter tails. As expected, we can see contrasting patterns between light- and heavy-tailed distributions. Median values, for Gamma and Uniform distributions, are equal to zero, regardless of sample length. In particular, box plots for Uniform distribution are completely concentrated in zero while the 75th percentile of Gamma box plots reaches 0.22 at most. This means that the biggest weight, or even the whole weight in case of $\omega = 0$, is assigned to a light-tailed kernel. In contrast, for the Generalized Pareto and left-truncated Cauchy distribution, box plots show higher values of ω , which lead to a mixture kernel where the heavy-tailed function has a

higher importance, with respect to the case of Gamma and Uniform distributions. Through the optimization of ω and h is possible to fit different tail behaviours without any a priori assumption. Nevertheless, results in Fig. 1 exhibit unsatisfactory performances, even with some differences, for most of the combinations. In the case of Gamma distribution the greatest biases are observed associated to return periods higher than 200 years and to the smallest sample lengths (25 and 50). For left-truncated Cauchy distributed samples, the extrapolation of quantiles related to return periods equal or greater than 200 years, provides high biases with respect to theoretical values, whatever the sample size. Performances for the Uniform distribution are instead comparable with those achieved through Hutson (2002). Finally, for the Generalized Pareto distribution results are very similar to those obtained for Gamma distribution for return periods up to 200 years. For higher return periods, instead, we observe a different behaviour

Fig. 4 Box plots of the optimized weight, ω , of the mixture kernel in KDE method

both with respect to the other distributions and methods. The median relative errors, in fact, assume values that move from positive to negative for increasing sample sizes.

5 Case study

Once more, we carried on a comparison among the investigated semi/nonparametric quantile estimators by applying them on a real precipitation dataset. We exploited a daily rainfall time series retrieved from the dataset ArCIS (Archivio Climatologico per l'Italia centro-Settentrionale, Climatological Archive for Central–Northern Italy) available at <https://www.arcis.it>. The database was chosen since data underwent a check for quality, time consistency, synchronicity, and statistical homogeneity (Pavan et al. 2019). Daily data, from January 1, 1961 to December 31, 2020 were aggregated on a monthly time scale and from this we selected only the accumulated rainfall depths related to the months of June, July, and August, to lead a seasonal analysis. We hence obtained three monthly rainfall values for each year, thus relying on 180 data points. As done concerning synthetic simulations, the case study analysis aims at providing a comparison with respect to the sample length and to the level of probability. We considered three sample lengths, $n = 25, 50, 75$ and we estimated quantiles $x(p)$ related to two probability levels, $p = 0.98, 0.99$, thereby investigating extrapolation of maxima. To provide a statistical significance we performed the analysis by using the random subsampling technique (Hartigan 1969). From the initial dataset, we randomly extracted 100 subsamples for each sample length n . Eventually, for all the 100 subsamples the quantiles were estimated. Whereas for subsamples with length equal to 25, both $x(p = 0.98)$ and $x(p = 0.99)$ can be extrapolated, for the remainder subsamples (with $n = 50$ and $n = 75$) only extrapolation of $x(p = 0.99)$ may be performed. As mentioned in Sec. 3, in hydrological practice precipitation depths are typically described by Gamma distribution. We hence considered to also include results gained through parametric extrapolation (PAR) from a Gamma distribution where parameters were estimated from each subset using maximum likelihood estimation. Concerning Scholz (1995) we disregarded SC7 method, since results reported in Sec. 4 reveal similar or even slightly worse performances compared to SC0. According to the simulation study we evaluated the relative error on estimates, η , referring to Equation (15). This time, though, the benchmark $x(p)$ was chosen to be the linear interpolation estimator falling within the range of the 180 rainfall observations and calculated with Equation (1). Results in Fig. 5 show boxplots of the 100 relative errors related to the three investigated semi/nonparametric extrapolation techniques and the

parametric extrapolation. Firstly, it can be easily noticed a discrepancy between parametric and semi/nonparametric methods. Relative errors from the parametric approach mostly exhibit positive values (estimated quantiles larger than interpolated ones) and the median relative errors appear not to approach zero with respect to the sample length, while dispersion slightly shrinks. Concerning the three semi/nonparametric methodologies, results are pretty consistent with synthetic simulation results. Median relative errors, whatever positive or negative, are always close to zero. They notably come even closer to zero with higher sample lengths. Similarly, errors IQR is reduced as the sample length increases. Focusing on sample length $n = 25$, the higher probability level increases uncertainty in the estimates, which is reflected by the higher dispersion of box plots. Best performances are gained by Hutson (2002) in terms of median error and by KDE in terms of dispersion. In the case of $n = 50$, Scholz (1995) provides the best results. Finally, for $n = 75$ Hutson (2002) and KDE slightly outperform Scholz (1995). Overall, the semi/nonparametric techniques gain quite similar and satisfying results.

6 Discussion

As already highlighted in the previous section, the performances of the presented methods strongly depend on the type of distribution. Therefore, considerations will be given individually for each distribution. Finally, general indications about the methods will be provided.

6.1 Left-truncated Cauchy distribution

The extrapolation from heavy-tailed distributions is often troublesome as also pointed out in Scholz and Tjoelker (1995), that observed errors twice as large for the Student t_5 distribution with respect to the other distributions considered, namely Uniform, reverse Weibull, Weibull, and Normal. The Cauchy distribution was excluded in the simulations carried out by Scholz and Tjoelker (1995) due to unrealistic estimations obtained for some of the generated samples and due to the high errors that made difficult a graphical comparison with the other distributions. Samples extracted from the Cauchy distribution are often characterized by the presence of a single value much greater than all the others and this may cause some of the problems encountered with this distribution. Also Hutson (2002) highlighted the poor performance of the proposed method in extrapolating heavy-tailed distributions.

Fig. 5 Box plots of the relative errors, η , of extreme quantiles extrapolated from the 100 subsamples of the daily rainfall dataset

6.2 Gamma distribution

Unlike Uniform, the Gamma distribution is asymmetric with an asymmetry that increases with the decreasing of the shape parameter. This may be the reason of the higher bias of KDE, that is known to have limitations with skewed distributions (Charpentier and Flachaire 2015; Kang et al. 2018). The Gamma distribution is also lower bounded, therefore attentions must be adopted in the application of KDE. Three alternatives to deal with positive data are here cited: (1) the choice of a kernel with a bounded support; (2) data transformation in order to match an infinite support of the kernel; (3) setting the estimated PDF to zero outside the admitted range and rescaling it so that the area under the curve remains one. In this case we disregarded the use of a bounded kernel, like the Gamma kernel (Chen 2000), since it would have not allowed a proper comparison between the three distributions. We tested both the second (results not reported) and third approach using a log transformation for the former and we selected the latter due to its best performance. As also pointed out by Zucchini (2003) the different procedures result in estimations that can be considerably different; hence, it may be important to test the uncertainty related with this choice when KDE is used with bounded data. Results from the case study, referred to $p = 0.98$ and $p = 0.99$, are typically in accordance to what is observed from synthetic simulations. Either way, Hutson (2002) mostly reports positive median errors and the highest IQRs. KDE, for such low probability levels, still shows moderate biases and small dispersions, as well as Scholz (1995).

6.3 Uniform distribution

Likewise the Gamma distribution, also the Uniform distribution requires to deal with the bounded support when KDE and Hutson (2002) are used. Regarding the former, the same procedure adopted for the Gamma distribution was chosen, while a logarithmic transformation of the data was applied for the latter (see Zucchini 2003). This passage implies that none of these two methods will produce extrapolated values outside the admitted range. Scholz (1995), on the contrary, can be equally applied to bounded or unbounded variables since the information regarding the upper bound is conserved in the value of the EVI. As a drawback, a wrong estimation of the EVI parameter may lead to extrapolated values greater than the real upper bound and this may explain the worst performance of this method for the Uniform distribution.

6.4 Generalized Pareto distribution

The Generalized Pareto distribution can be of three types depending on the value of the shape parameter γ : moving from an exponential ($\gamma = 0$, light-tailed) to a Pareto ($\gamma > 0$, heavy-tailed) to a Beta ($\gamma < 0$). The value of EVI is equal to the shape parameter. In this study we adopted a set of parameters that was fitted to discharge data and for which the shape parameter resulted equal to 0.15. This means that we are studying a heavy-tailed distribution, even though with a lighter tail than the one of left-truncated Cauchy distribution, characterized by $\text{EVI} = 1$. The performances of the selected methods, in fact, are more similar to the ones of Gamma distribution rather than left-truncated Cauchy distribution with similar median errors but higher IQRs, confirming the increasing difficulty in extrapolating

from heavy-tailed distributions. In order to apply KDE, as done also for Gamma and left-truncated Cauchy distribution, it is necessary to deal with the lower bound of data when applying a kernel with infinite support. The choice done to correct this may influence the resulting performances.

6.5 Limitations and improvements

The performance of Scholz (1995), as well as KDE, depends on the estimation of some parameters, in particular the EVI and k for the former and h and ω for the latter. The variability of the estimation of the EVI around the true value may be the reason of the higher IQR of this method with respect to the others and it may also explain the reduction of IQR with increasing sample sizes. We recall that Scholz (1995) assumes that EVT holds therefore it is reasonable to expect a better performance with increasing sample size. Nevertheless, different distributions converge to their domain of attraction with different rates. Hence, a larger sample size may be required to reach the true value of the EVI in case of distributions with lower convergence rates. Figure 3 shows that also a sample length of 500 is still too short to estimate correctly the EVI. We point out that Scholz (1995) can be applied also with different estimators of the EVI and different criteria for the choice of the k -greatest order statistics.

Even though the choice of the kernel function is usually considered of less importance than bandwidth selection (Chen 2000; Wilks 2011), in the present comparison we observed that this is not true when the extrapolation capacity is considered, since tail properties of KDE depend on the type of kernel (Lall et al. 1993). The use of a mixture kernel avoids reverting to the problem already present in parametric methods for which results depend on the distribution function chosen. In the present comparison we proposed an application of KDE that can adapt to a broad range of circumstances without assuming any prior knowledge of how data are distributed. Nevertheless, the kernel function as well as the bandwidth may be chosen also considering the data at hand. What we want to stress out is that the kernel choice is not unimportant as is also confirmed by the different median values of ω obtained for the four distributions. A further improvement of KDE may be obtained combining a variable bandwidth with the mixture kernel (Wang et al. 2020).

Unlike KDE and Scholz (1995), Hutson (2002) does not require the choice or estimation of any parameter. Even though this may yield to less flexibility, the method has the advantage to be simple and easy to use and implement, while still providing competitive results when light-tailed distributions are of interest.

Although in this work we dealt only with quantile estimators, the method of Hutson (2002) was developed in order to improve the construction of confidence intervals based on bootstrap replications, therefore providing a useful instrument when we are interested in quantifying the uncertainty of the estimation. Also Scholz (1995) can be used to estimate confidence intervals using a different value of α in Eq. (4). Nevertheless, the theoretical basis that justifies a linear extrapolation does not hold when α is different from 0.5. Regardless, Scholz and Tjoelker (1995) observed a good fit using linear extrapolation also for $\alpha \neq 0.5$ while they discouraged the use of a quadratic extrapolation.

Analysis revealed that performances of methods vary depending on the behaviour of extreme values and in particular on the tail weight. When there are no a priori information on the parametric distribution which could describe data nor on their asymptotic distribution, still it is possible to get information on the tail weight, by resorting to sample estimates of tail weight indices as reported in Rosenberger and Gasko (1983). Such indices could thus help in the choice of the best semi/nonparametric method.

7 Conclusions

The estimate of extreme quantiles is a key issue in hydrological applications. In this study, the performances among three semi/nonparametric techniques for extrapolation of extreme quantiles are compared, namely Hutson (2002), KDE, and Scholz (1995). Hutson (2002) has only few applications in the hydrological context. KDE was widely exploited in flood frequency analysis, however we enhanced the method by the use of a new flexible type of kernel (Wang et al. 2020). At our knowledge this kernel was never applied before. On the other hand, Scholz (1995) is here implemented for hydrological purposes for the first time. The analysis, based on synthetic series, does not indicate that one technique typically outperforms the others. Methods' performances vary with data distribution, sample length, and level of probability or return period. It is therefore recommended to consciously exploit nonparametric methods taking into account the type of dataset. This work can be seen as a support in the choice of the most suitable technique even if it cannot provide an exhaustive analysis of all methodologies available in literature. Firstly, when dealing with extrapolation, it is crucial to be aware of observed variables support. In fact, data transformation is sometimes required, as in the case of KDE and Hutson (2002) for bounded samples. Data transformation methods must be carefully managed since they could lead to further biases in quantiles extrapolation. Whether it is possible to make some assumptions on

observed data distribution or on their asymptotic distribution, or simply on the tail weight (e.g., with the use of tail indices, see Rosenberger and Gasko 1983), some specific guidelines can be furnished relying on our analysis. KDE poorly performs when applied to skewed data or heavy-tailed distributions. In the former case it would be recommendable to use a skewed kernel function as well. Hutson (2002) generates large biases in the case of heavy-tailed distributions. On the other hand, Scholz (1995) provides the best performances for heavy-tailed distributions, when observed data have a size greater than 50. Whereas information on data distribution is not available, we suggest to prefer Scholz (1995) when sample lengths are greater than 75, whatever the return period. We indeed observed that such a sample size is generally suitable for a

reliable estimation of the EVI, for any type of distribution. More challenging is the extrapolation of quantiles when sample length is shorter than 75: Scholz (1995) leads to wrong estimates of EVI and consequently to large biases in quantile estimates. In these cases, Hutson (2002) or KDE should be favoured.

Appendix

See Figs. 6, 7, 8, 9, 10.

Fig. 6 Box plots of relative errors, η , of extreme quantiles extrapolated from samples with a length equal to 25

Fig. 7 Box plots of relative errors, η , of extreme quantiles extrapolated from samples with a length equal to 75

Fig. 8 Box plots of relative errors, η , of extreme quantiles extrapolated from samples with a length equal to 100

Fig. 9 Box plots of relative errors, η , of extreme quantiles extrapolated from samples with a length equal to 200

Fig. 10 Box plots of relative errors, η , of extreme quantiles extrapolated from samples with a length equal to 500

Acknowledgements We wish to acknowledge the support by the Italian Ministry of University and Research through the PRIN2017 project RELaid (<https://relaid.wordpress.com/>). We thank prof. Fritz W. Scholz for some personal communications and providing us with some of his works. We would also like to thank the Associate Editor Dr. Francesco Serinaldi and the two anonymous Reviewers for their useful suggestions and remarks.

Funding Open access funding provided by Politecnico di Milano within the CRUI-CARE Agreement.

Availability of data and material The data used for this analysis can be freely downloaded from <http://doi.org/10.5281/zenodo.4756673>.

Code availability The R and Python codes used to produce results and figures can be freely downloaded from <http://doi.org/10.5281/zenodo.4756673>.

Declarations

Conflict of interest All authors declare that they have no conflict of interest.

Consent for publication All authors read the manuscript and agreed to publication.

Consent to participate All authors have consented to participate.

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

- Adamowski K (1985) Nonparametric kernel estimation of flood frequencies. *Water Resour Res* 21(11):1585–1590. <https://doi.org/10.1029/WR021i011p01585>
- Adamowski K (1989) A monte carlo comparison of parametric and nonparametric estimation of flood frequencies. *J of Hydrol* 108:295–308. [https://doi.org/10.1016/0022-1694\(89\)90290-4](https://doi.org/10.1016/0022-1694(89)90290-4)
- Adamowski K, Feluch W (1990) Nonparametric flood-frequency analysis with historical information. *J of Hydraul Eng* 116(8):1035–1047. [https://doi.org/10.1061/\(ASCE\)0733-9429\(1990\)116:8\(1035\)](https://doi.org/10.1061/(ASCE)0733-9429(1990)116:8(1035))
- Alila Y, Mtroui A (2002) Implications of heterogeneous flood-frequency distributions on traditional stream-discharge prediction techniques. *Hydrol Process* 16(5):1065–1084. <https://doi.org/10.1002/hyp.346>
- Apipattanas S, Rajagopalan B, Lall U (2010) Local polynomial-based flood frequency estimator for mixed population. *J of Hydrol Eng* 15(9):680–691. [https://doi.org/10.1061/\(ASCE\)HE.1943-5584.0000242](https://doi.org/10.1061/(ASCE)HE.1943-5584.0000242)
- Bardsley WE (1989) Using historical data in nonparametric flood estimation. *J Hydrol* 108:249–255. [https://doi.org/10.1016/0022-1694\(89\)90286-2](https://doi.org/10.1016/0022-1694(89)90286-2)
- Barth NA, Villarini G, White K (2019) Accounting for mixed populations in flood frequency analysis: bulletin 17c perspective. *J Hydrol Eng* 24(3):04019002. [https://doi.org/10.1061/\(ASCE\)HE.1943-5584.0001762](https://doi.org/10.1061/(ASCE)HE.1943-5584.0001762)
- Beirlant J, Goegebeur Y, Segers J, Teugels JL (2004) *Statistics of extremes: theory and applications*. John Wiley and Sons, West Sussex, United Kingdom
- Beirlant J, Dierckx G, Guillou A (2005) Estimation of the extreme-value index and generalized quantile plots. *Bernoulli* 11(6):949–970. <https://doi.org/10.3150/bj/1137421635>
- Bezak N, Rusjan S, Fijavž MK, Mikoš M, Šraj M (2017) Estimation of suspended sediment loads using copula functions. *Water* 9(8):628. <https://doi.org/10.3390/w9080628>
- Bezak N, Zabret K, Šraj M (2018) Application of copula functions for rainfall interception modelling. *Water* 10(8):995. <https://doi.org/10.3390/w10080995>
- Bezak N, Jemec Auflič M, Mikoš M (2019) Application of hydrological modelling for temporal prediction of rainfall-

- induced shallow landslides. *Landslides* 16(7):1273–1283. <https://doi.org/10.1007/s10346-019-01169-9>
- Brazdil R, Kundzewicz ZW (2006) Historical hydrology. *Hydrol Sci J* 51(5):733–738. <https://doi.org/10.1623/hysj.51.5.733>
- Brennan B, Marc M, Yan S (2017) Predicting utah ground snow loads with prism. *J Struct Eng* 143(9):04017126. [https://doi.org/10.1061/\(ASCE\)ST.1943-541X.0001870](https://doi.org/10.1061/(ASCE)ST.1943-541X.0001870)
- Carreau J, Naveau P, Sauquet E (2009) A statistical rainfall-runoff mixture model with heavy-tailed components. *Water Resour Resh.* <https://doi.org/10.1029/2009WR007880>
- Charpentier A, Flachaire E (2015) Log-transform kernel density estimation of income distribution. *L'Actual Econ* 91(1–2):141–159. <https://doi.org/10.7202/1036917ar>
- Chen SX (2000) Probability density function estimation using gamma kernels. *Ann Inst Stat Math* 52(3):471–480. <https://doi.org/10.1023/A:1004165218295>
- Cheng MY, Sun S (2006) Bandwidth selection for kernel quantile estimation. *J Chin Stat Assoc* 44(3):271–295
- Chow VT, Maidment DR, Mays LW (1988) *Applied hydrology*. Civil Engineering Series, McGraw Hill, New York
- De Michele C (2019) Advances in deriving the exact distribution of maximum annual daily precipitation. *Water* 11(11):2322. <https://doi.org/10.3390/w11112322>
- De Michele C, Avanzi F (2018) Superstatistical distribution of daily precipitation extremes: a worldwide assessment. *Sci Rep* 8(1):1–11. <https://doi.org/10.1038/s41598-018-31838-z>
- Dekkers ALM, Einmahl JH, De Haan L (1989) A moment estimator for the index of an extreme-value distribution. *Ann Stat* 17(4):1833–1855. <https://www.jstor.org/stable/2241667>
- Duin RPW (1976) On the choice of smoothing parameters for parzen estimators of probability density functions. *IEEE Trans Comput* 25(11):1175–1179
- El Adlouni S, Bobée B, Ouarda TBMJ (2008) On the tails of extreme event distributions in hydrology. *J Hydrol* 355(1):16–33. <https://doi.org/10.1016/j.jhydrol.2008.02.011>
- El Fels AEA, Alaa N, Bachnou A, Rachidi S (2018) Flood frequency analysis and generation of flood hazard indicator maps in a semi-arid environment, case of ourika watershed (western high atlas, morocco). *J Afr Earth Sci* 141:94–106. <https://doi.org/10.1016/j.jafrearsci.2018.02.004>
- Engeland K, Hisdal H, Frigessi A (2004) Practical extreme value modelling of hydrological floods and droughts: A case study. *Extremes* 7(1):5–30. <https://doi.org/10.1007/s10687-004-4727-5>
- England JFJ, Cohn TA, Faber BA, Stedinger JR, Jr WOT, Veilleux AG, Kiang JE, R JMR (2019) Guidelines for determining flood flow frequency—bulletin 17c (ver. 1.1, may 2019). Tech. Rep. book 4, chap. B5, U.S. Geological Survey Techniques and Methods, <https://doi.org/10.3133/tm4B5>, <http://pubs.er.usgs.gov/publication/tm4B5>
- Epanechnikov VA (1969) Non-parametric estimation of a multivariate probability density. *Theory Probab Appl* 14(1):153–158. <https://doi.org/10.1137/1114019>
- Evin G, Merleau J, Perreault L (2011) Two-component mixtures of normal, gamma, and gumbel distributions for hydrological applications. *Water Resour Res.* <https://doi.org/10.1029/2010WR010266>
- Ghorbani MA, Ruskeepaa H, Singh VP, Sivakumar B (2011) Flood frequency analysis using mathematica. *Turk J Eng Env Sci* 34(3):171–188
- Hartigan JA (1969) Using subsample values as typical values. *J Am Stat Assoc* 64(328):1303–1317. <https://doi.org/10.1080/01621459.1969.10501057>
- Hill BM (1975) A simple general approach to inference about the tail of a distribution. *Ann Stat* 3(5):1163–1174. <https://www.jstor.org/stable/2958370>
- Hutson AD (2000) A composite quantile function estimator with applications in bootstrapping. *J Appl Stat* 27(5):567–577. <https://doi.org/10.1080/02664760050076407>
- Hutson AD (2002) A semi-parametric quantile function estimator for use in bootstrap estimation procedures. *Stat Comput* 12(4):331–338. <https://doi.org/10.1023/A:1020783911574>
- Kang YJ, Noh Y, Lim OK (2018) Kernel density estimation with bounded data. *Struct Multidiscip Optim* 57(1):95–113. <https://doi.org/10.1007/s00158-017-1873-3>
- Kim K, Heo JH (2002) Comparative study of flood quantiles estimation by nonparametric models. *J Hydrol* 260(1–4):176–193. [https://doi.org/10.1016/S0022-1694\(01\)00613-8](https://doi.org/10.1016/S0022-1694(01)00613-8)
- Kim T, Valdés JB, Yoo C (2003) Nonparametric approach for estimating return periods of droughts in arid regions. *J Hydrol Eng* 8(5):237–246. [https://doi.org/10.1061/\(ASCE\)1084-0699\(2003\)8:5\(237\)](https://doi.org/10.1061/(ASCE)1084-0699(2003)8:5(237))
- Lall U (1995) Recent advances in nonparametric function estimation: hydrologic applications. *Rev Geophys* 33(S2):1093–1102. <https://doi.org/10.1029/95RG0034310.1029/95RG00343>
- Lall U, Rajagopalan B (2016) *Nonparametric methods*. McGraw-Hill Education, New York, NY, United States of America, pp 25–1–25–5. *Handbook of applied hydrology*
- Lall U, Moon Y, Bosworth K (1993) Kernel flood frequency estimators: bandwidth selection and kernel choice. *Water Resour Res* 29(4):1003–1015. <https://doi.org/10.1029/92WR02466>
- Lee H, Kang K (2015) Interpolation of missing precipitation data using kernel estimations for hydrologic modeling. *Adv Meteorol* 2015:1–12. <https://doi.org/10.1155/2015/935868>
- Lekina A, Chebana F, Ouarda TBMJ (2014) Weighted estimate of extreme quantile: an application to the estimation of high flood return periods. *Stoch Environ Res Assess* 28(2):147–165. <https://doi.org/10.1007/s00477-013-0705-2>
- MacDonald A, Scarrott C, Lee D, Darlow B, Reale M, Russell G (2011) A flexible extreme value mixture model. *Comput Stat Data Anal* 55(6):2137–2157. <https://doi.org/10.1016/j.csda.2011.01.005>
- Martinez-Villalobos C, Neelin JD (2019) Why do precipitation intensities tend to follow gamma distributions? *J Atmos Sci* 76(11):3611–3631. <https://doi.org/10.1175/JAS-D-18-0343.1>
- McKee TB, Doesken NJ, Kleist J (1993) The relationship of drought frequency and duration to time scales. Paper presented at the 8th Conference on Applied Climatology, American Meteorological Society, Anaheim, CA, United States of America
- Moon YI, Lall U (1994) Kernel quantile function estimator for flood frequency analysis. *Water Resour Res* 30(11):3095–3103. <https://doi.org/10.1029/94WR01217>
- Moon YI, Lall U, Bosworth K (1993) A comparison of tail probability estimators for flood frequency analysis. *J Hydrol* 151(2–4):343–363. [https://doi.org/10.1016/0022-1694\(93\)90242-2](https://doi.org/10.1016/0022-1694(93)90242-2)
- Nashwan MS, Ismail T, Ahmed K (2018) Flood susceptibility assessment in kelantan river basin using copula. *Int J Eng Technol* 7(2):584–590
- Parzen E (1979) Nonparametric statistical data modeling. *J Am Stat Assoc* 74(365):105–121
- Pavan V, Antolini G, Barbiero R, Berni N, Brunier F, Cacciamani C, Cagnati A, Cazzuli O, Cicogna A, De Luigi C et al (2019) High resolution climate precipitation analysis for north-central Italy, 1961–2015. *Clim Dyn* 52(5):3435–3453. <https://doi.org/10.1007/s00382-018-4337-6>
- Pickands J (1975) Statistical inference using extreme order statistics. *Ann Stat* 3(1):119–131. <https://doi.org/10.1214/aos/1176343003>
- Rosenberger JL, Gasko M (1983) Comparing location estimators: trimmed means, medians, and trimean. In: Hoaglin DC,

- Mosteller F, Tukey JW (eds) Understanding robust and explanatory data analysis. John Wiley and Sons, New York, pp 297–338
- Salvadori G, De Michele C, Kotegoda NT, Rosso R (2007) Extremes in nature: an approach using copulas, vol 56. Springer, Dordrecht, The Netherlands
- Scholz FW (1995) Nonparametric tail extrapolation. Tech. Rep. ISSTECH-95-014, Boeing Information and Support Services, Seattle, WA, United States of America
- Scholz FW, Tjoelker R (1995) Nonparametric tail extrapolation: Simulation results. Tech. Rep. ISSTECH-95-015, Boeing Information and Support Services, Seattle, WA, United States of America
- Serinaldi F (2009) Assessing the applicability of fractional order statistics for computing confidence intervals for extreme quantiles. *J Hydrol* 376(3–4):528–541. <https://doi.org/10.1016/j.jhydrol.2009.07.065>
- Serinaldi F (2011) Analytical confidence intervals for index flow flow duration curves. *Water Resour Res*. <https://doi.org/10.1029/2010WR009408>
- Serinaldi F, Grimaldi S, Abdolhosseini M, Corona P, Cimini D (2012) Testing copula regression against benchmark models for point and interval estimation of tree wood volume in beech stands. *Eur J Res* 131(5):1313–1326. <https://doi.org/10.1007/s10342-012-0600-2>
- Sheather SJ (2004) Density estimation. *Stat Sci* 19(4):588–597. <https://doi.org/10.1214/088342304000000297>
- Sheather SJ, Jones MC (1991) A reliable data-based bandwidth selection method for kernel density estimation. *J R Stat Soc Ser B (Methodol)* 53(3):683–690. <https://doi.org/10.1111/j.2517-6161.1991.tb01857.x>
- Silverman BW (1986) Density estimation for statistics and data analysis, vol 26. CRC Press, Boca Raton
- Singh VP (1998) Entropy-based parameter estimation in hydrology. Kluwer Academic Publishers, London. <https://doi.org/10.1007/978-94-017-1431-0>
- Sol'áková T, De Michele C, Vezzoli R (2014) Comparison between parametric and nonparametric approaches for the calculation of two drought indices: SPI and SSI. *J Hydrol Eng* 19(9):04014010. [https://doi.org/10.1061/\(ASCE\)HE.1943-5584.0000942](https://doi.org/10.1061/(ASCE)HE.1943-5584.0000942)
- Tencaliec P, Favre AC, Naveau P, Prieur C, Nicolet G (2020) Flexible semiparametric generalized pareto modeling of the entire range of rainfall amount. *Environmetrics* 31(2):2582. <https://doi.org/10.1002/env.2582>
- UNDRR (2019) Global assessment report on disaster risk reduction. Tech. rep, United Nations Office for Disaster Risk Reduction (UNDRR), Geneva, Switzerland
- Volpi E (2019) On return period and probability of failure in hydrology. *WIREs Water* 6(3):1340. <https://doi.org/10.1002/wat2.1340>
- Wand MP, Jones MC (1994) Kernel smoothing. CRC Press, Boca Raton
- Wang S, Li A, Wen K, Wu X (2020) Robust kernels for kernel density estimation. *Econ Lett* 191:109138. <https://doi.org/10.1016/j.econlet.2020.109138>
- Waylen P, Woo M (1982) Prediction of annual floods generated by mixed processes. *Water Resour Res* 18(4):1283–1286. <https://doi.org/10.1029/WR018i004p01283>
- Wei L, Wang D, Hutson AD (2015) An investigation of quantile function estimators relative to quantile confidence interval coverage. *Commun Stat Theory Methods* 44(10):2107–2135. <https://doi.org/10.1080/03610926.2013.775304>
- Wilks D (2011) Statistical methods in the atmospheric sciences, vol 100. Academic press, San Diego, California
- Young DS, Mathew T (2014) Improved nonparametric tolerance intervals based on interpolated and extrapolated order statistics. *J Nonparametr Stat* 26(3):415–432. <https://doi.org/10.1080/10485252.2014.906594>
- Zucchini W (2003) Applied smoothing techniques Part 1: kernel density estimation. Temple University, Philadelphia

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