AGU PUBLICATIONS

Water Resources Research

RESEARCH ARTICLE
10.1002/2015WR018502

Key Points:
- Critical review of representative methods for GP (generalized Pareto) threshold detection
- Application to 1714 overcentennial daily rainfall records from the NOAA-NCDC database
- For daily rainfall, GP threshold estimates range between 2 and 12 mm/d with a mean value of 6.5 mm/d

Correspondence to:
A. Langousis,
andlag@alum.mit.edu

Citation:
Langousis, A., A. Mamalakis, M. Puliga, and R. Deidda (2016), Threshold detection for the generalized Pareto distribution: Review of representative methods and application to the NOAA NCDC daily rainfall database, *Water Resour. Res.*, *52*, doi:10.1002/2015WR018502.

Received 12 DEC 2015
Accepted 12 FEB 2016
Accepted article online 15 FEB 2016

Threshold detection for the generalized Pareto distribution: Review of representative methods and application to the NOAA NCDC daily rainfall database

Andreas Langousis[1], Antonios Mamalakis[1], Michelangelo Puliga[2], and Roberto Deidda[3]

[1]Department of Civil Engineering, University of Patras, Patras, Greece, [2]IMT Institute for Advanced Studies Lucca, Lucca, Italy, [3]Dipartimento di Ingegneria Civile, Ambientale e Architettura, University of Cagliari, Cagliari, Italy

Abstract In extreme excess modeling, one fits a generalized Pareto (GP) distribution to rainfall excesses above a properly selected threshold *u*. The latter is generally determined using various approaches, such as nonparametric methods that are intended to locate the changing point between extreme and nonextreme regions of the data, graphical methods where one studies the dependence of GP-related metrics on the threshold level *u*, and Goodness-of-Fit (GoF) metrics that, for a certain level of significance, locate the lowest threshold *u* that a GP distribution model is applicable. Here we review representative methods for GP threshold detection, discuss fundamental differences in their theoretical bases, and apply them to 1714 overcentennial daily rainfall records from the NOAA-NCDC database. We find that nonparametric methods are generally not reliable, while methods that are based on GP asymptotic properties lead to unrealistically high threshold and shape parameter estimates. The latter is justified by theoretical arguments, and it is especially the case in rainfall applications, where the shape parameter of the GP distribution is low; i.e., on the order of 0.1–0.2. Better performance is demonstrated by graphical methods and GoF metrics that rely on preasymptotic properties of the GP distribution. For daily rainfall, we find that GP threshold estimates range between 2 and 12 mm/d with a mean value of 6.5 mm/d, while the existence of quantization in the empirical records, as well as variations in their size, constitute the two most important factors that may significantly affect the accuracy of the obtained results.

1. Introduction

Due to its importance in quantifying hydrologic risk, statistical estimation of extreme rainfall has been a pressing and widely studied problem in engineering hydrology [see e.g., *Bernard*, 1932; *Demarée*, 1985; *Coles and Tawn*, 1996; *Koutsoyiannis et al*., 1998; *Katz et al*., 2002; *Koutsoyiannis*, 2004a, 2004b; *Cooley et al*., 2007; *Veneziano et al*., 2006, 2007, 2009; *Veneziano and Yoon*, 2013; *Deidda*, 2010; *El Adlouni and Ouarda*, 2010; *Papalexiou and Koutsoyiannis*, 2013; *Papalexiou et al*., 2013; *Serinaldi and Kilsby*, 2014, among others]. While several efforts have been devoted to modeling rainfall extremes using scaling representations of rainfall and their corresponding linkages to large deviation (LD) theory [see e.g., *Schertzer and Lovejoy*, 1987; *Hubert et al*., 1993; *Lovejoy and Schertzer*, 1995; *Deidda et al*., 1999, 2004, 2006; *Deidda*, 2000; *Veneziano and Langousis*, 2005; *Veneziano et al*., 2006, 2009; *Langousis and Veneziano*, 2007; *Langousis et al*., 2009, 2013; *Veneziano and Lepore*, 2012; *Veneziano and Yoon*, 2013, and the review in *Veneziano and Langousis*, 2010], the statistical analysis of annual maxima series (AMS) and partial duration series (PDS) still remains the main pillar for practical estimation of hydrologic extremes from empirical records (i.e., under preasymptotic conditions). To obtain the AMS from empirical observations, one extracts the block-maxima over different years [see e.g., *Bernard*, 1932; *Demarée*, 1985; *Cunnane*, 1973; *Madsen et al*., 1997a,1997b; *Koutsoyiannis et al*., 1998; *Katz et al*., 2002; *Papalexiou and Koutsoyiannis*, 2013; *Serinaldi and Kilsby*, 2014], while PDS are constructed by retaining all excesses above a sufficiently high threshold *u* (thus, this last approach is also referred to as peaks over threshold, POT) [see e.g., *Davison and Smith*, 1990; *Wang*, 1991; *Rosbjerg and Madsen*, 1992, 1995; *Madsen et al*., 1997a, 1997b, 2002; *Lang et al*., 1999; *Willems*, 2000; *Deidda and Puliga*, 2006; *Cooley et al*., 2007; *Veneziano et al*., 2007; *Deidda*, 2010; *Serinaldi and Kilsby*, 2014].

The wide use of AMS and PDS for the estimation of hydrologic extremes stems from extreme value (EV) and extreme excess (EE) theories, respectively. According to EV theory, if under suitable normalization the

cumulative distribution function (CDF) of the maximum $M_n = \max\{X_1, X_2, X_3, \ldots, X_n\}$ of n independent copies of random variable X converges to a nondegenerate distribution $G_{\max}$ as $n \to \infty$, then this distribution should have the generalized extreme value (GEV) form [see e.g., *Fisher and Tippett*, 1928; *Gnedenko*, 1943; *Jenkinson*, 1955, and more recently *Coles*, 2001, and *Lucarini et al*., 2016]:

$$G_{\max}(y) = \exp\left\{-\left[1+k\left(\frac{y-\psi}{\lambda}\right)\right]^{-1/k}\right\} \tag{1}$$

where k, λ, and ψ are the shape, scale, and location parameters of the distribution, respectively.

Under a similar setting, extreme excess (EE) theory suggests that if the CDF $F_{X_u}(x) = \{F_X(x+u) - F_X(u)\}/\{1 - F_X(u)\}$ of the scaled excesses $X_u = [X - u | X > u]$ of random variable X above threshold u converges to a nondegenerate distribution Q_u as u increases and $F_X(u) \to 1$, then this distribution should have the generalized Pareto (GP) form [see e.g., *Balkema and de Haan*, 1974; *Pickands*, 1975; *Leadbetter et al*., 1983, and more recently *Coles*, 2001, and *Lucarini et al*., 2016]:

$$Q_u(y) = P[X - u \leq y | X > u] = 1 - \left(1 + \xi\frac{y}{a_u}\right)^{-1/\xi} \tag{2}$$

where, similar to equation (1), ξ and a_u are the shape and scale parameters of the distribution, respectively, with the latter depending on the threshold u used to determine the excesses $Y = X - u$ of random variable X.

The specific form of the distributions in equations (1) and (2) depend on the upper tail of the distribution of random variable X (usually referred to as the parent distribution); see e.g., *Gumbel* [1958], and more recently *Coles* [2001]. For k, $\xi = 0$, equations (1) and (2) reduce to the Gumbel (i.e., EV type 1, $F(y) = \exp\{-\exp(-(y-\psi)/\lambda)\}$) and exponential (i.e., $F(y) = 1 - \exp(-y/a_u)$) forms, respectively; for k, $\xi > 0$, the corresponding distributions have heavy upper tails that behave like power functions with exponents $-1/k$ and $-1/\xi$, respectively; whereas for k, $\xi < 0$, the distributions in equations (1) and (2) exhibit finite upper bounds. For the GEV case, $k > 0$ and $k < 0$ correspond to the so-called Frechet and Weibull (i.e., EV 2 and 3) distribution forms, respectively.

Clearly, the GP and GEV families have important theoretical connections. As shown by *Pickands* [1975], when the threshold u increases and $F_X(u) \to 1$: (1) the distribution of the scaled excesses $X_u = [X - u | X > u]$ converges to a nondegenerate distribution Q_u if and only if the maximum of $n \to \infty$ independent copies of X is attracted to a nondegenerate distribution $G_{\max}$, (2) Q_u has the generalized Pareto form in equation (2) with the same shape parameter as $G_{\max}$ in equation (1) (i.e., $\xi = k$), and (3) the location and scale parameters of Q_u and $G_{\max}$ are theoretically linked through [see e.g., *Coles*, 2001, p. 77]:

$$a_u = \lambda + \xi(u - \psi) \tag{3}$$

Although the AMS and PDS approaches should lead asymptotically to the same results, one expects that the PDS approach is more suitable for extreme quantile estimations from empirical records than the AMS approach. For example, when implementing the AMS approach using continuous rainfall data, one retains solely the block (e.g., yearly) maxima discarding a large portion of the available hydrologic information. This makes the estimated distribution parameters to exhibit significant variability and be sensitive to outliers [see e.g., *Coles and Tawn*, 1996; *Coles et al*., 2003; *Koutsoyiannis*, 2004b; *Deidda*, 2010; *Langousis et al*., 2013]. To that extent, several studies dedicated significant efforts in quantifying the performance of AMS and PDS approaches.

Cunnane [1973] was the first to compare the two methods, using an exponential distribution model (i.e., a GP model with shape parameter $\xi = 0$) for the excesses of flood magnitudes, and a Gumbel distribution (i.e., a GEV model with shape parameter $k = 0$) for yearly flood maxima. The analysis concluded that the PDS approach leads to quantile estimates with smaller variance, if the length of the PDS exceeds $1.65N$, where N is the number of years in record. Using Monte Carlo techniques, *Madsen et al*. [1997a] generalized the findings of *Cunnane* [1973] to the most general case of GP and GEV distributions (i.e., shape parameters ξ, $k \neq 0$), and showed that the relative performance of the two approaches depends on the parameter estimation method used (i.e., probability weighted moments, maximum likelihood, and method of moments) and the value of the shape parameter. The study concluded that for heavy-tailed distributions (i.e., ξ, $k > 0$), commonly met in hydrologic applications, the PDS approach with a GP distribution model (PDS/GP)

performs better than the AMS approach with a GEV distribution model (AMS/GEV), independent of the estimation method used. Similar findings have been reported by *Wang* [1991], *Madsen et al.* [1997b], and *Tanaka and Takara* [2002] among others.

Since the PDS/GP approach is more efficient for hydrologic applications, several studies have focused on developing new, or improving existing, methods for fitting a GP distribution model to data. For example, *Hosking and Wallis* [1987] derived the equations for GP parameter estimation using probability weighted moments (PWM) and, for the case when $|\xi| < 0.5$, compared the obtained estimates to those of maximum likelihood (ML) and simple moments (SM). The study concluded that ML estimation, although asymptotically unbiased, does not display its competitive advantage for sample sizes smaller than 500, SM is generally reliable except for the case when $\xi > 0.2$, while PWM estimation has better overall performance for heavy-tailed distributions (i.e., $\xi > 0$) commonly met in hydrologic applications.

Improvements to the classical ML and SM methods have been proposed by *Grimshaw* [1993], *Dupuis* [1998], *Dupuis and Tsao* [1998], *Martins and Stedinger* [2001a, 2001b], *Juarez and Schucany* [2004], and *Zhang* [2007] among others; for a detailed review, see *Rosbjerg and Madsen* [2004] and *Deidda and Puliga* [2009]. For example, *Grimshaw* [1993] used a univariate transformation of the ML equations to develop a computationally efficient approach to fit a GP distribution model to data, while *Dupuis* [1998] used a weighting scheme to introduce a generalization of the ML equations, referred to as optimal bias-robust estimator (OBRE) [see also *Huber*, 1964; *Hampel*, 1968, 1974; *Hampel et al.*, 1986]. *Dupuis and Tsao* [1998] introduced a hybrid estimator based on SM and PWM, which incorporates a simple auxiliary constraint on the feasibility of the estimates, and *Martins and Stedinger* [2001a, 2001b] introduced a robust version of the classical ML estimator for small samples (referred to as Generalized Maximum Likelihood approach), which uses Bayesian priors to restrict the estimated parameters within reasonable ranges. Generalizations of the ML equations for GP parameter estimation have been suggested, also, by *Juarez and Schucany* [2004] and *Zhang* [2007].

While several advancements have been suggested for fitting a GP distribution model to data, application of the PDS/GP approach requires prior determination of the threshold u. The latter is a nontrivial issue that significantly affects the validity of the obtained estimates [see e.g., *Rosbjerg and Madsen*, 1992; *Lang et al.*, 1999; *Coles*, 2001; *Deidda and Puliga*, 2006; *Tancredi et al.*, 2006; *Deidda*, 2010; *Chavez-Demoulin and Davison*, 2012; *Scarrott and MacDonald*, 2012] and cannot be easily tackled through standard inference techniques. For high thresholds and, thus, small sample sizes, the obtained estimates are characterized by small bias and increased variance [see e.g., *Coles*, 2001], whereas for small thresholds, the obtained estimates can be significantly biased due to deviations of the empirical distribution from a perfect GP model [see e.g., *Adamowski*, 2000].

In order to detect a proper threshold u to extract PDS from data and fit a GP distribution model, several statistical methods have been proposed (see e.g., *Scarrott and MacDonald* [2012] for an extended review), which can be grouped into three broad categories: (1) nonparametric methods that are intended to locate the changing point between extreme and nonextreme regions of the data [see e.g., *Gerstengarbe and Werner*, 1989, 1991; *Werner and Gerstengarbe*, 1997; *Domonkos and Piotrowicz*, 1998; *Lasch et al.*, 1999; *Cebriàn et al.*, 2003; *Cebriàn and Abaurrea*, 2006; *Karpouzos et al.*, 2010, among others], (2) graphical methods where one searches for linear behavior of the GP parameters (or related metrics) with increasing threshold level [see e.g., *Davison and Smith*, 1990; *Lang et al.*, 1999; *Tanaka and Takara*, 2002; *Cebriàn et al.*, 2003; *Deidda*, 2010; *Scarrott and MacDonald*, 2012; *Das and Ghosh*, 2013, among others], and (3) Goodness-of-Fit (GoF) tests that, for any given level of significance, locate the lowest threshold u that a GP distribution model is applicable. The latter category of methods can be subdivided in two groups: (a) statistical tests that use some type of quadratic error metric to quantify the departure between the empirical and fitted distributions [see e.g., *Stephens*, 1986; *Ahmad et al.*, 1988; *Dupuis*, 1998; *Choulakian and Stephens*, 2001; *Laio*, 2004; *Deidda and Puliga*, 2006; *Deidda* 2007; *Serinaldi and Kilsby*, 2014], and (b) Hill-assumption-based methods, which detect the lowest threshold u that the log-transformed data exhibit some type of (asymptotic) exponential behavior [see e.g., *Beirlant et al.*, 1996, 2006; *Goegebeur et al.*, 2008, among others].

The variety of existing methods for GP threshold detection, the fundamental differences in their theoretical bases, and their relative performance when dealing with different types of data (e.g., heavy-tailed versus light-tailed distributions, and continuous versus quantized samples) make threshold detection an open question that can be addressed solely on the basis of a specific application. For example, two recent studies by *Deidda and Puliga* [2006, 2009] showed that independent of the GP parameter estimation method used

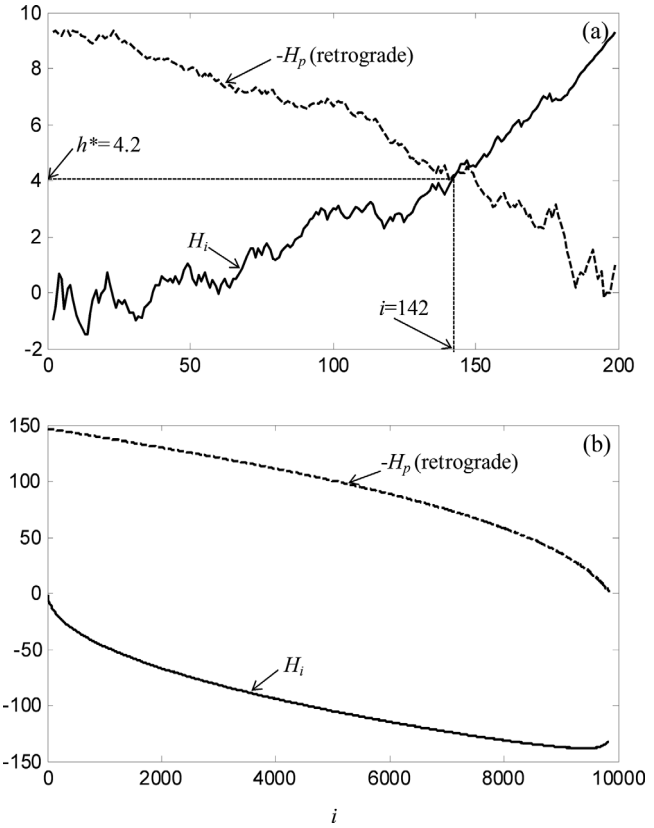

Figure 1. Application of Gerstengarbe and Werner (GW) plot to: (a) 200 synthetic realizations drawn from a GP distribution with threshold $u = 0$, scale parameter $a_0 := a_{u=0} = 15$ and shape parameter $\xi = 0.1$; (b) the positive rainrates extracted from a 126 year record of daily rainfall observations from Australia.

(ML, SM, PWM etc.), application of GoF tests to quantized (i.e., rounded off) data may lead to significantly biased and highly uncertain estimates of distribution parameters.

In an effort to bridge this gap, this study reviews representative methods for GP threshold detection, discusses fundamental differences in their theoretical bases, uses theoretical arguments and numerical simulations to study their appropriateness for fitting a GP distribution model to finite samples (as is the case of empirical observations), and applies them to 1714 overcentennial daily rainfall records from the NOAA-NCDC (National Oceanic and Atmospheric Administration-National Climatic Data Center) database. To the best of our knowledge, this is the first time that a detailed intercomparison study of GP threshold detection methods is presented, followed by an extended application to rainfall records collected worldwide.

In section 2, we start by reviewing the theoretical basis of each category of methods and discuss their limitations. Section 3 presents details on the NOAA-NCDC rainfall data set, and discusses the results obtained when applying the GP threshold detection methods to empirical rainfall records. Our findings show that in the case of rainfall, where the shape parameter ξ is low (i.e., on the order of 0.1–0.2, see sections 2.4 and 3.3 and Appendix B for a theoretical investigation), threshold detection methods based on GP asymptotics underperform relative to other methods, leading to significant biases. Graphical and GoF methods based on GP preasymptotic properties are more reliable and also lead to similar results, with the former category being less sensitive to data quantization and variations of the sample size. Section 4 discusses the main findings of this work, and points toward future research directions.

2. GP Threshold Detection Methods

2.1. A Nonparametric Method: Gerstengarbe and Werner (GW) Plot

Gerstengarbe and Werner [1989] used the sequential version of Mann-Kendal (MK) test [see e.g., *Sneyers*, 1963, 1975; *Maasch*, 1988; *Gerstengarbe and Werner*, 1999; *Karpouzos et al.*, 2010] to devise a nonparametric method, referred to as Gerstengarbe and Werner (GW) plot, that is intended to detect the starting point of the extreme region of a sample. The latter can be used as the threshold u to extract the PDS from data [see e.g., *Gerstengarbe and Werner*, 1989; *Werner and Gerstengarbe*, 1997; *Domonkos and Piotrowicz*, 1998; *Lasch et al.*, 1999; *Cebrìan et al.*, 2003; *Cebrìan and Abaurrea*, 2006]. In what follows, we review the basic assumptions of the GW method, and discuss its theoretical limitations.

Define the series of differences $\Delta_i = X_{i+1,n} - X_{i,n}$, $i = 1, \ldots, n-1$, of the (ascending) order statistics $X_{1,n} \le X_{2,n} \le \ldots \le X_{n,n}$ in a sample of size n (i.e., Δ_i, $i = 1, \ldots, n-1$, is the series of differences between adjacent values in the sorted sample). In the case when random variable X is not upper bounded, it follows from probability theory that as one moves further into the upper tail of the distribution of X and i increases, the

expected value $E[\Delta_i]$ (corresponding to the magnitude of the differences), also increases. Based on this theoretical result, *Gerstengarbe and Werner* [1989] suggested to detect threshold u (i.e., the starting point of extreme behavior in a sample) as the point x^* where series $\{\Delta_i, i = 1, \ldots, n-1\}$ exhibit a statistically significant change. For the latter purpose, *Gerstengarbe and Werner* [1989] defined variable $H_i' := \sum_{k=1}^{i} n_k$ ($i = 1, \ldots, n-1$), where n_k is the rank of value Δ_k in the series of the differences $\{\Delta_j, j = 1, \ldots, k\}$, and applied the sequential version of the MK test to check the null hypothesis that the differences $\{\Delta_j, j = 1, \ldots, i\}$ are independent and identically distributed (iid) and, therefore, no change in the data exists. Under the null hypothesis, n_k follows a uniform distribution in the interval $[1, k]$ (i.e., $U[1, k]$), $E[n_k] = \frac{k-1}{2}$, $\mathrm{Var}[n_k] = \frac{k^2-1}{12}$ and, hence, $E[H_i'] = \sum_{k=1}^{i} E[n_k] = \frac{i(i-1)}{4}$, $\mathrm{Var}[H_i'] = \sum_{k=1}^{i} \mathrm{Var}[n_k] = \frac{i(i-1)(2i+5)}{72}$. For large values of i (say $i > 10$) [see e.g., *Kendall*, 1938; *Mann*, 1945], the standardized random variable: $H_i = (H_i' - b_i)/\sqrt{c_i}$, where $b_i = \frac{i(i-1)}{4}$ and $c_i = \frac{i(i-1)(2i+5)}{72}$, can be approximated by a standard normal distribution.

Following the sequential MK test procedure [see e.g., *Sneyers*, 1963, 1975; *Maasch*, 1988, and more recently *Karpouzos et al*., 2010], *Gerstengarbe and Werner* [1989] approximate the starting point of change in the series of differences $\{\Delta_i, i = 1, \ldots, n-1\}$ (i.e., the starting point of extreme behavior), as the sample value that corresponds to the intersection point h^* of series H_i and the retrograde of series $-H_p$ (see Figure 1a and discussion below). H_p is calculated by applying the same procedure used to obtain series H_i, to the series of differences from end to start; i.e., $\{\Delta_p, p = 1, \ldots, n-1\} \equiv \{\Delta_i, i = n-1, \ldots, 1\}$. The level of significance that the null hypothesis (i.e., of no change in the series of differences) is rejected equals the probability $1 - P[|H_i| \le h^*]$.

Although the idea of detecting the extreme behavior in a sample based on the statistics of the differences of the sorted series seems logical, application of the sequential version of the MK test to check the null hypothesis of no change in the data exhibits an important theoretical limitation: unless the observations x_i, $i = 1, \ldots, n$ are uniformly distributed (i.e., they are described by a constant probability density function, PDF) the differences $\Delta_i = X_{i+1,n} - X_{i,n}$, $i = 1, \ldots, n-1$, of the order statistics $X_{1,n} \le X_{2,n} \le \ldots \le X_{n,n}$, are by definition not identically distributed. Thus, the null hypothesis of randomness in the series of differences $\{\Delta_i, i = 1, \ldots, n-1\}$, tested using the sequential version of the MK test, is rejected on the basis of probability theory for any sample drawn from a nonuniform distribution. This theoretical result applies to the whole sample, and not just beyond a specific point that denotes transition to an extreme region. Consequently, the procedure suggested by *Gerstengarbe and Werner* [1989] to detect the starting point of extreme behavior in a sample lacks statistical meaning and, therefore, its application should be avoided.

To illustrate the above arguments, Figure 1a shows the H_i and the retrograde of $-H_p$ series obtained from a synthetic sample of size $n = 200$, drawn from a GP distribution (see equation (2)) with threshold $u = 0$, scale parameter $a_0 := a_{u=0} = 15$ and shape parameter $\xi = 0.1$. One sees that intersection of the two curves occurs at $i = 142$, which corresponds to threshold $u = 18.8$. The obtained value significantly overestimates the GP threshold $u = 0$, which according to EE theory denotes transition to extreme behavior.

In the case of quantized data (i.e., a common case in rainfall applications [see e.g., *Deidda and Puliga*, 2006, 2009; *Deidda*, 2007, 2010, and references therein]), the GW method leads to strange results; see Figure 1b and discussion below. More precisely, independent of the level of data quantization, the series of differences $\{\Delta_i, i = 1, \ldots, n-1\}$ calculated from the sorted sample incorporate large chunks of zeros, which preclude series H_i and $-H_p$ from intersecting. As an example, Figure 1b illustrates an application of the GW method to the positive rainrates of a 126 year record of daily rainfall observations in Australia, obtained from NOAA-NCDC rainfall database (station ID name: ASN00021043; see section 3.1 for details on the data). Although the quantization of rainfall data is small (i.e., on the order of 0.1 mm/d; see section 2.3 and discussion in section 3.1), it suffices to prevent the two H-curves from intersecting, making the GW method inapplicable.

It follows from the discussion above that the nonparametric method of *Gerstengarbe and Werner* [1989], which has been used to detect the starting point of the extreme region of a sample [see e.g., *Gerstengarbe and Werner*, 1989, 1991; *Werner and Gerstengarbe*, 1997; *Domonkos and Piotrowicz*, 1998; *Lasch et al*., 1999; *Cebrià n et al*., 2003; *Cebrià n and Abaurrea*, 2006, among others], is theoretically incorrect and may lead to

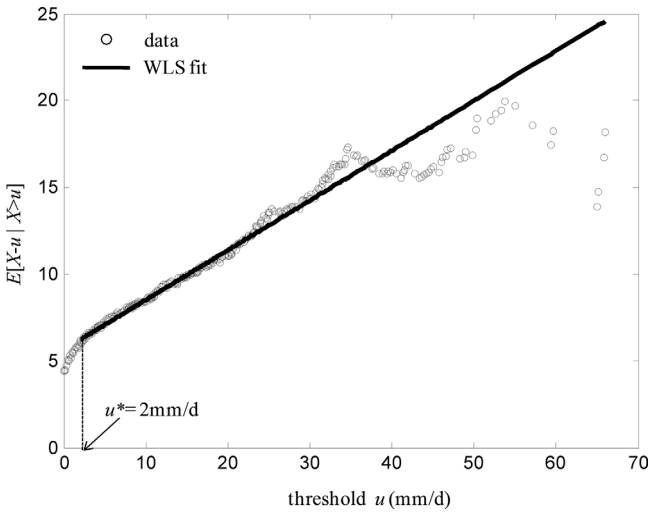

Figure 2. Application of the mean residual life plot (MRLP) method, using the stepwise procedure described in section 2.2, to the positive rainrates of the 126 year rainfall record of Figure 1b.

spurious results, especially in the case of quantized samples. For these reasons, its extended application to rainfall records from the NOAA-NCDC database (see section 3) is meaningless and, therefore, skipped.

2.2. Graphical Methods: Mean Residual Life Plot

The most popular graphical method for threshold estimation is based on the mean residual life plot (MRLP) [see e.g., *Davison and Smith*, 1990; *Lang et al*., 1999; *Coles*, 2001; *Cebrià n et al*., 2003; *Scarrott and MacDonald*, 2012; *Das and Ghosh*, 2013, and references therein]. As shown in Appendix A, if the scaled excesses $X_{u^*} = [X - u^* | X > u^*]$ of random variable X above threshold u^* are GP distributed then, for any threshold $u \geq u^*$, the scaled excesses $X_u = [X - u | X > u]$ are also GP distributed with the same shape parameter ξ, scale parameter a_u that depends linearly on u:

$$a_u = a_{u^*} + \xi(u - u^*) \tag{4}$$

and mean value:

$$e(u) = E[X - u | X > u] = \frac{a_u}{1-\xi} = \frac{a_{u^*} + \xi(u - u^*)}{1-\xi} = Au + B \tag{5}$$

where a_u and a_{u^*} are the scale parameters of the distribution when fitted to the excesses above thresholds u and u^*, respectively, and $A = \xi/(1-\xi)$, $B = (a_{u^*} - \xi u^*)/(1-\xi)$ are the slope and intercept of the linear relation.

Based on equation (5), one can obtain a proper threshold u^* to extract PDS from data in order to fit a GP distribution model, by: (a) plotting the mean of the excesses $e(u)$ as a function of the threshold u (referred to as mean residual life plot), and (b) identifying the lowest threshold u^* above which $e(u)$ increases linearly with u. By doing so, one keeps the sample size sufficiently large to reduce the statistical variability of the obtained results, while minimizing deviations of the distribution of the excesses from a perfect GP model. As an example, Figure 2 shows the MRLP applied to the positive rainrates of the historical rainfall record used in Figure 1b. One sees that above thresholds u on the order of 2–2.5 mm/d, $e(u)$ depends linearly on u, indicating that the scaled excesses can be effectively approximated by a GP distribution model. For values of u larger than 25–30 mm/d, the sample size of the excesses reduces significantly, leading to a considerable increase of the estimation variance.

Despite its simplicity and being particularly suited for fitting a GP distribution model to data, an apparent limitation of MRLP is that threshold detection is conducted on the basis of subjective criteria; i.e., usually through visual inspection of the expected linear behavior [see e.g., *Coles*, 2001; *Scarrott and MacDonald*, 2012]. This limitation (common to all graphical methods) becomes important when the number of series to be analyzed is large, as it is the case in this study, where we consider more than 1700 daily rainfall records from the NOAA-NCDC database; see section 3.1. In an effort to automate threshold detection, we applied the procedure outlined below:

a. We estimated the mean value of the excesses $e(u) = E[X - u | X > u]$ above different thresholds $u_i = X_{i,n}$, $i = 1, 2, \ldots, n - 10$ (see points in Figure 2), where $X_{i,n}$ denotes the ith (ascending) order statistic in a

sample of size n (selection of $X_{n-10,n}$ as maximum threshold ensures a minimum of 10 excesses to calculate the conditional mean $e(u)$).

b. For each u_i ($i = 1, 2, \ldots, n - 20$) from (a), we used the method of weighted least squares (WLS) to fit a linear model to all points (u_j, $e(u_j)$) that satisfy $j \geq i$ (selection of $u_{n\text{-}20}$ as maximum threshold ensures a minimum of 10 conditional means, i.e., $e(u)$ points, for the linear regression). To account for the increase of the estimation variance of $e(u)$ with increasing threshold u, the weight w_j applied to each point (u_j, $e(u_j)$) was taken to be inversely proportional to the variance of $e(u)$, assuming independence of the excesses. In this case, $w_j = (n - j)/\mathrm{Var}[X - u_j | X > u_j]$.

c. We determined the threshold u^* as the lowest threshold u_i ($i = 1, 2, \ldots, n - 20$) which corresponds to a local minimum of the weighted mean square error (WMSE) function of the linear regression.

After threshold u^* is determined, estimates of the shape ξ and scale a_{u^*} parameters of the distribution can be obtained using the slope A and intercept B of the linear fit:

$$\xi = A/(1 + A), a_{u^*} = B(1-\xi) + \xi u^* \tag{6}$$

For the case presented in Figure 2, the aforementioned procedure resulted in $u^* = 2$ mm/d, $A = 0.287$, $B = 5.632$ mm/d, and using equation (6), one obtains $\xi = 0.223$ and $a_{u^*} = 4.822$ mm/d. Similar values for the GP distribution parameters (i.e., $\xi = 0.226$ and $a_{u^*} = 4.776$ mm/d) were obtained by applying the maximum likelihood (ML) method [see e.g., *Hosking and Wallis*, 1987; *Grimshaw*, 1993] to the excesses above threshold $u^* = 2$ mm/d, suggesting that the resulting PDS can be effectively approximated using a GP model (see also Figure 6 and discussion in section 3.3).

2.3. Goodness-of-Fit (GoF)-Based Methods

In an effort to maximize the retained information when extracting PDS from river flow records, *Choulakian and Stephens* [2000, 2001] suggested the use of goodness-of-fit statistics to identify the lowest threshold u above which the excesses can effectively be approximated by a GP distribution model. In their method, referred to as "failure-to-reject", one simultaneously increases the threshold used to extract PDS from data (e.g., starting from the smallest nonzero sample value), until the null hypothesis (H_0) of GP distributed excesses is not rejected at a desired significance level. To do so, *Choulakian and Stephens* [2001] used the W^2 Crámer-von Mises and A^2 Anderson-Darling statistics [*Anderson and Darling*, 1952, 1954]:

$$W^2 = \frac{1}{12n} + \sum_{i=1}^{n} \left[F(X_{i,n}) - \frac{2i-1}{2n} \right]^2 \tag{7}$$

$$A^2 = -n - \sum_{i=1}^{n} \frac{2i-1}{n} (\log[F(X_{i,n})] + \log[1 - F(X_{n+1-i,n})]) \tag{8}$$

where $X_{1,n} \leq X_{2,n} \leq \ldots \leq X_{n,n}$ are (ascending) order statistics, and F is the theoretical CDF tested for fitting.

Both W^2 and A^2 quantify the deviations between a selected theoretical distribution model, F, and the empirical CDF, $F_{emp}^{(n)}$, calculated from a sample of size n, and belong to the Crámer-von Mises family of quadratic statistics [see e.g., *Anderson and Darling*, 1952, 1954; *Stephens*, 1986; *Ahmad et al*., 1988; *Deidda and Puliga*, 2006, and references therein]:

$$S^2 = n \int_{-\infty}^{+\infty} \left[F_{emp}^{(n)}(x) - F(x) \right]^2 \psi(x) dF(x) \tag{9}$$

where $\psi(x)$ is a weight function. W^2 is obtained by setting $\psi(x) = 1$ and weights equally the whole distribution, while A^2 corresponds to the case when $\psi(x) = \{F(x)[1 - F(x)]\}^{-1}$. Hence A^2 assigns larger weights to observations located in the (upper or lower) distribution tail.

Choulakian and Stephens [2000, 2001] used theoretical arguments and Monte Carlo simulations to study the distributions of W^2 and A^2 statistics in equations (7) and (8) for samples drawn from a GP distribution, and for the case when either one or both distribution parameters (i.e., scale and shape) are estimated using the method of maximum likelihood (ML). Based on theoretical arguments, they concluded that asymptotically as $n \to \infty$, the distributions of W^2 and A^2 are invariant under changes of the GP scale parameter and, further, produced tables of the quantiles of the asymptotic distributions for different values of the shape

Table 1. Quantiles of the Crámer-von Mises (W^2) Statistic in Equation (7), Calculated Using 10,000 Synthetic Realizations of Samples With Sizes $n = 500$, 5000, and 10,000, Drawn From a GP Distribution With Threshold $u = 0$, Scale Parameter $a_0 := a_{u=0} = 10$, Different Values of the Shape Parameter $\xi = 0$, 0.1, 0.2, 0.3, and Levels of Quantization $\Delta = 0$ (Continuous Sample), 0.1, 0.5, and 1

	W^2									
	90% Quantiles			95% Quantiles			99% Quantiles			
Shape ξ	$n = 500$	$n = 5000$	$n = 10{,}000$	$n = 500$	$n = 5000$	$n = 10{,}000$	$n = 500$	$n = 5000$	$n = 10{,}000$	Data Type
0	0.121	0.121	0.122	0.150	0.151	0.152	0.220	0.221	0.222	Continuous ($\Delta = 0$)
0.1	0.115	0.116	0.116	0.143	0.143	0.143	0.209	0.210	0.209	
0.2	0.112	0.111	0.111	0.137	0.137	0.137	0.200	0.199	0.199	
0.3	0.107	0.107	0.108	0.132	0.132	0.132	0.191	0.191	0.191	
0	0.128	0.165	0.207	0.158	0.199	0.249	0.229	0.288	0.334	Quantized $\Delta = 0.1$ ($\Delta/a_0 = 0.01$)
0.1	0.118	0.156	0.196	0.143	0.187	0.233	0.211	0.262	0.323	
0.2	0.116	0.151	0.186	0.141	0.181	0.220	0.201	0.264	0.306	
0.3	0.110	0.143	0.179	0.137	0.173	0.215	0.202	0.242	0.291	
0	0.218	1.000	1.817	0.256	1.074	1.915	0.349	1.231	2.099	Quantized $\Delta = 0.5$ ($\Delta/a_0 = 0.05$)
0.1	0.212	0.947	1.718	0.248	1.016	1.805	0.343	1.169	1.994	
0.2	0.199	0.895	1.625	0.236	0.968	1.705	0.323	1.109	1.869	
0.3	0.189	0.852	1.550	0.223	0.913	1.633	0.295	1.053	1.803	
0	0.476	3.284	6.274	0.533	3.412	6.434	0.653	3.659	6.765	Quantized $\Delta = 1$ ($\Delta/a_0 = 0.1$)
0.1	0.447	3.072	5.881	0.500	3.182	6.034	0.606	3.414	6.326	
0.2	0.423	2.891	5.531	0.471	2.991	5.670	0.578	3.208	5.961	
0.3	0.399	2.736	5.226	0.448	2.835	5.361	0.551	3.043	5.624	

parameter ξ. For finite samples with size $n \geq 25$, they concluded that the asymptotic quantiles are very close to what one obtains using Monte Carlo simulations.

In a more recent study, *Deidda and Puliga* [2006] used Monte Carlo simulations to study how the asymptotic distributions of W^2 and A^2 are influenced by the parameter estimation method used (i.e., maximum likelihood, ML; simple moments, SM; probability weighted moments, PWM) and, more importantly, the level of quantization in the data. In accordance with the findings of *Hosking and Wallis* [1987], their results show that for values of the shape parameter $\xi \geq 0.2$, SM underperforms relative to the other two methods (i.e., due to moment divergence). In addition, the study concluded that contrary to the case of continuous samples where the asymptotic distributions of W^2 and A^2 are invariant under scale changes, in the case of quantized samples, the asymptotic distributions of W^2 and A^2 depend on the ratio Δ/a_u, where Δ is the level of data quantization, and a_u is the scale parameter of the GP model; see equation (2).

As an illustration, Tables 1 and 2 show quantiles of W^2 and A^2 statistics at different probability levels, calculated using 10,000 synthetic realizations of samples with sizes $n = 500$, 5000, and 10,000, drawn from a GP distribution with threshold $u = 0$, scale parameter $a_0 := a_{u=0} = 10$, different values of the shape parameter

Table 2. Same as Table 1, but for the Anderson-Darling (A^2) Statistic in Equation (8)

	A^2									
	90% Quantiles			95% Quantiles			99% Quantiles			
Shape ξ	$n = 500$	$n = 5000$	$n = 10{,}000$	$n = 500$	$n = 5000$	$n = 10{,}000$	$n = 500$	$n = 5000$	$n = 10{,}000$	Data Type
0	0.793	0.794	0.799	0.971	0.969	0.977	1.403	1.397	1.415	Continuous ($\Delta = 0$)
0.1	0.763	0.766	0.766	0.933	0.937	0.934	1.346	1.349	1.338	
0.2	0.743	0.742	0.741	0.905	0.904	0.901	1.299	1.295	1.297	
0.3	0.719	0.719	0.720	0.876	0.874	0.871	1.257	1.248	1.245	
0	0.870	1.596	2.408	1.058	1.857	2.715	1.523	2.470	3.324	Quantized $\Delta = 0.1$ ($\Delta/a_0 = 0.01$)
0.1	0.803	1.557	2.334	0.965	1.792	2.636	1.376	2.312	3.276	
0.2	0.799	1.535	2.278	0.965	1.757	2.530	1.356	2.283	3.132	
0.3	0.773	1.481	2.232	0.947	1.701	2.498	1.336	2.267	3.082	
0	1.962	11.973	22.706	2.197	12.490	23.423	2.787	13.535	24.794	Quantized $\Delta = 0.5$ ($\Delta/a_0 = 0.05$)
0.1	1.907	11.634	21.958	2.150	12.111	22.604	2.766	13.130	23.871	
0.2	1.830	11.247	21.283	2.090	11.782	21.906	2.620	12.784	23.119	
0.3	1.768	10.889	20.678	1.983	11.370	21.257	2.475	12.322	22.516	
0	4.264	33.684	65.469	4.596	34.459	66.495	5.354	35.985	68.526	Quantized $\Delta = 1$ ($\Delta/a_0 = 0.1$)
0.1	4.090	32.229	62.716	4.414	32.973	63.743	5.017	34.375	65.679	
0.2	3.937	30.918	60.254	4.231	31.612	61.204	4.849	32.995	63.211	
0.3	3.789	29.825	58.014	4.068	30.515	58.941	4.694	31.871	60.591	

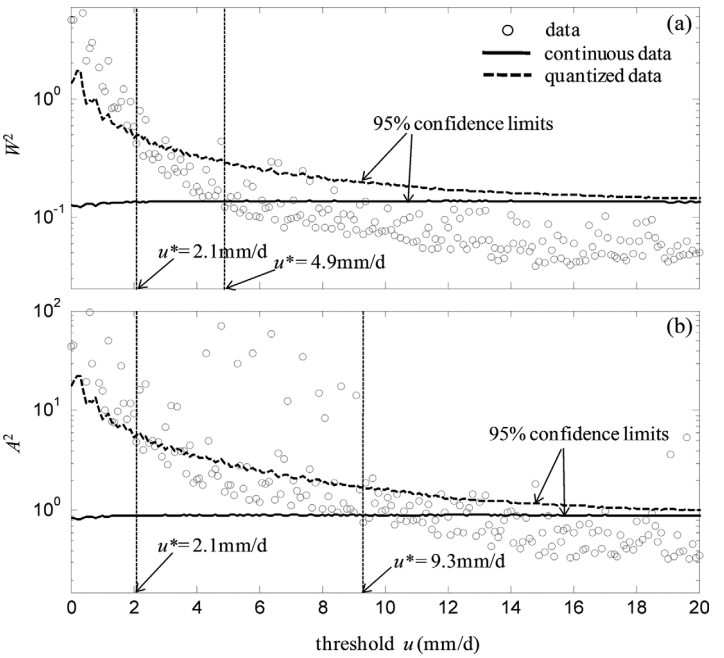

Figure 3. Application of the "failure-to-reject" method to the positive rainrates of Figure 1b, using: (a) the W^2 Crámer-von Mises statistic (equation (7)), and (b) the A^2 Anderson-Darling statistic (equation (8)). In both subplots, the solid black lines denote the 95% quantiles of the corresponding statistics as obtained from *Choulakian and Stephens* [2001, Table 2] (i.e., asymptotic conditions and continuous data). The dashed black lines denote the 95% quantiles under preasymptotic conditions and in the presence of data quantization, calculated through Monte Carlo simulations.

$\xi = 0$, 0.1, 0.2, 0.3, and levels of quantization $\Delta = 0$ (continuous sample), 0.1, 0.5, and 1. In our calculations, both shape and scale parameters of the GP distribution were considered unknown and estimated using the method of ML [*Grimshaw*, 1993]. An observation one makes is that, in the case of continuous samples, the calculated W^2 and A^2 quantiles are insensitive to the sample size n (i.e., denoting fast convergence of the distributions of W^2 and A^2 to those under asymptotic conditions, as $n \to \infty$), whereas in the case of quantized data, the corresponding quantiles increase with increasing sample length n and quantization ratio Δ/a_0. The observed behavior should be attributed to the fact that, for any nonzero quantization level, as the sample size increases the empirical CDF (i.e., having the form of a step function, due to quantization) is densified nonuniformly at prespecified discrete locations along the observation axis, with decreasing number of densification points as the quantization level Δ increases. Under this setting, it follows from equations (7) and (8) that an increase of Δ/a_0 for fixed sample size n, or an increase of the sample size n for fixed Δ/a_0, would necessarily lead to an increase of the calculated W^2 and A^2 quantiles at any given probability level. As a consequence, in the case of heavily quantized data, no limiting distributions for W^2 and A^2 should exist.

Based on the aforementioned findings, and since quantization is quite common in rainfall records [see e.g., *Deidda and Puliga*, 2006, 2009; *Deidda*, 2007, 2010, and references therein], one concludes that application of the "failure-to-reject" method using the asymptotic distributions of W^2 and A^2 for continuous data should lead to an overestimation of the proper threshold u^* to extract PDS; see below and section 3.

As an example, Figures 3a and 3b present estimates of W^2 and A^2 in equations (7) and (8), respectively, as a function of threshold u, for the positive rainrates of the historical rainfall record used in Figure 1b. The solid lines denote the 95% quantiles of W^2 and A^2, obtained from *Choulakian and Stephens* [2001, Table 2] (i.e., asymptotic conditions and continuous data), as a function of the shape parameter, ξ, estimated using the ML method for each threshold u. Note that use of ML for the estimation of the GP shape parameter is compulsory, as the quantiles reported in *Choulakian and Stephens* [2001, Table 2] have been obtained using ML. The dashed lines denote the 95% quantiles of W^2 and A^2 under preasymptotic conditions and in the presence of data quantization observed in the historical record. To obtain the dashed lines, for each threshold u, we calculated the corresponding quantiles of W^2 and A^2 using 10,000 synthetic realizations of GP distributed samples of length n_u equal to that of the PDS extracted from the historical rainfall record, and parameters (a_u, ξ) estimated by applying the method of maximum likelihood to the extracted series. To match the quantization mixture observed in the historical rainfall record, 77.8% of the simulated values were rounded at quantization level $\Delta = 0.1$ mm/d, and the remaining 22.2% at $\Delta = 0.2$ mm/d. The aforementioned estimates were obtained by applying the rounding-off rule estimator (RRE) developed by *Deidda* [2007] to the historical record of rainfall observations, to obtain the fractions of rainfall measurements that can be grouped into different categories of data quantization levels.

One sees that while the solid lines (i.e., referring to continuous data) are approximately horizontal (i.e., small variations are due to the variability of ξ estimates), the dashed lines (i.e., accounting for data quantization) decrease fast with increasing threshold level, as larger threshold values lead to reduced PDS lengths; see also Tables 1 and 2. To better evaluate the effect of data quantization on GP threshold detection, note that when neglecting data quantization (i.e., solid lines) the calculated thresholds at the 95% confidence level are 4.9 mm/d for W^2 (see Figure 3a) and 9.3 mm/d for A^2 (see Figure 3b), whereas when accounting for quantization in the historical data (i.e., dashed lines) both statistics lead to threshold estimates of 2.1 mm/d, similar to the MRLP method (see Figure 2 and section 2.2).

An extended application of the "failure-to-reject" method to NOAA-NCDC rainfall dataset is presented in section 3, where we compare the results of GP threshold detection under asymptotic conditions (i.e., as $n \to \infty$) assuming continuous data (e.g., solid lines in Figures 3a and 3b), and in the presence of data quantization found in the historical samples (e.g., dashed lines in Figures 3a and 3b).

2.4. Hill-Assumption-Based Methods

Let Y be a Pareto-type random variable with CDF [see e.g., *Beirlant et al*., 2004, 2006; *de Haan and Ferreira*, 2006; *Goegebeur et al*., 2008]:

$$F_Y(y) = 1 - y^{-1/\xi} l_F(y), y > 0 \tag{10}$$

where $\xi > 0$ is the shape parameter of the distribution ($1/\xi$ is also referred to as Pareto tail index), and l_F is a slowly varying function at infinity that satisfies:

$$\lim_{y\to\infty} \frac{l_F(\lambda y)}{l_F(y)} = 1, \text{ for any } \lambda > 0 \tag{11}$$

In the specific case of the GP distribution in equation (2),

$$l_F(y) = (1/y + \xi/a_u)^{-1/\xi} \tag{12}$$

It follows from equation (10) that under asymptotic conditions as $y \to \infty$, $1 - F_Y(y) \propto y^{-1/\xi}$ and, hence, a log-transformed Pareto-type variable is attracted to an exponential distribution.

Based on this asymptotic link between Pareto type and exponential distributions, *Hill* [1975] [see also *Beirlant et al*., 1996, 1999; *Vandewalle et al*., 2007] suggested the following estimator (usually referred to as Hill estimator) for the shape parameter ξ:

$$\hat{\xi}_{m,n} = \frac{1}{m}\sum_{j=1}^{m}\left[\log\left(X_{n-j+1,n}\right) - \log\left(X_{n-m,n}\right)\right] = \frac{1}{m}\sum_{j=1}^{m} j\left[\log\left(X_{n-j+1,n}\right) - \log\left(X_{n-j,n}\right)\right] \tag{13}$$

where $X_{1,n} \le X_{2,n} \le \ldots \le X_{n,n}$ are ascending order statistics in a sample of size n, and m may take integer values from 1 to $n - 1$, denoting the number of the largest values in the sample used for estimation. The linkage between equation (13) and the aforementioned asymptotic result becomes more evident if one views $\hat{\xi}_{m,n}$ as an estimator of the mean excess function $e_{\log X}(\log u^*) = E[\log X - \log u^* | X > u^*]$ of the log-transformed data above threshold $u^* = X_{n-m,n}$: the $m + 1$ largest value in a sample of size n.

Strictly speaking, the estimator in equation (13) is unbiased only asymptotically as $m, n \to \infty$ and $m/n \to 0$ [see e.g., *Goegebeur et al*., 2008]. However, *Beirlant et al*. [1996] [see also *Kratz and Resnick*, 1996; *Beirlant et al*., 1999] suggested its use also under preasymptotic conditions: within the range (i.e., the m largest values in a sample) that the complementary CDF (CCDF) of the data displays a log-log linear behavior. In this case, detection of the proper threshold u^* to extract PDS from data can be done by: (a) plotting $-\log[j/(n+1)]$ (i.e., the negatively signed logarithmically transformed empirical exceedance probability of the jth largest value in a sample, using a Weibull plotting-position formula) against $\log[X_{n-j+1,n}]$, for $j = 1, 2, 3, \ldots, n$ (usually referred to as Pareto quantile plot) [see e.g., *Beirlant et al*., 1996, 1999; *Kratz and Resnick*, 1996, among others], and (b) setting $u^* = X_{n-m,n}$: the $n - m$ ascending order statistic beyond which approximate linearity of the Pareto quantile plot holds. The latter is usually quantified on the basis of subjective criteria, including visual inspection [see e.g., *Hall*, 1990; *Hall and Welsh*, 1985; *Resnick and Stărică*, 1997; *Scarrott and MacDonald*, 2012], or addressed as a diagnostic regression problem using WLS [see e.g., *Beirlant et al*., 1996, 1999; *Kratz and Resnick*, 1996].

In an effort to statistically delimit the region of approximate linearity in Pareto quantile plots for tail index estimation, *Goegebeur et al*. [2008] adjusted two well-known goodness-of-fit (GoF) kernel statistics (devised to test exponentiallity of empirical samples), namely the Jackson [*Jackson*, 1967] and Lewis [*Lewis*, 1965] kernel statistics. Their derivations were based on the aforementioned asymptotic link between log-transformed Pareto-type and exponentially distributed variables, leading to:

$$J_{m,n}=\sqrt{m}\frac{\frac{1}{m}\sum_{j=1}^{m}K_J\left(\frac{j}{m+1}\right)Z_j}{\hat{\xi}_{m,n}},\ L_{m,n}=\sqrt{m}\frac{\frac{1}{m}\sum_{j=1}^{m}K_L\left(\frac{j}{m+1}\right)Z_j}{\hat{\xi}_{m,n}} \tag{14}$$

where $J_{m,n}$ and $L_{m,n}$ are the Jackson and Lewis modified kernel statistics, respectively, calculated using the m largest values in a sample of size n, $K_J(\theta)=-1-\log(\theta)$ and $K_L(\theta)=\theta-0.5$ are appropriate kernel functions [see *Beirlant et al*., 1996], $Z_j=j[\log(X_{n-j+1,n})-\log(X_{n-j,n})]$, and $\hat{\xi}_{m,n}$ is Hill's estimator in equation (13). Given the asymptotic normality of $\hat{\xi}_{m,n}$ at the limit as $m, n\to\infty$ and $m/n\to 0$ [see e.g., *Hall*, 1982; *Davis and Resnick*, 1984; *Csörgő et al*., 1985; *Csörgő and Mason*, 1985; *Haeusler and Teugels*, 1985; *Beirlant and Teugels*, 1987; *Beirlant et al*., 1996], the modified Jackson and Lewis kernel statistics in equation (14) are also normally distributed with (approximately) zero mean (depending on the particular form of function l_F) [see *Goegebeur et al*., 2008] and variance 1 and 1/12, respectively.

To identify the optimal threshold u^* to extract the sample fraction (i.e., the PDS) for tail estimation, one simultaneously increases the threshold used to extract PDS from data (this is equivalent to decreasing the number of values m used to estimate the slope in a Pareto quantile plot), until the null hypothesis (H_0) of exponentially distributed excesses of the log-transformed sample (suggesting linearity of the Pareto quantile plot) is not rejected at a desired significance level.

While the study of *Goegebeur et al*. [2008] sheds light on tail index estimation under asymptotic conditions, use of asymptotic arguments under preasymptotic conditions may influence the obtained results. This is especially the case when convergence of the limit in equation (11) is slow [see *Beirlant et al*., 1999, 2004; *Scarrott and MacDonald*, 2012]. More precisely, the modified Jackson and Lewis kernel statistics do not test the GP assumption (as the Crámer-von Mises and Anderson-Darling statistics presented in section 2.3), but rather the null hypothesis of linearity in a Pareto quantile plot above some high threshold value. Since exponentiality of a log-transformed GP variable is established only asymptotically as one moves into the upper tail of the distribution, one concludes that the modified Jackson and Lewis kernel statistics should lead to significantly higher thresholds than those estimated using methods based on GP distribution properties valid, also, under preasymptotic conditions (see sections 2.2 and 2.3). As shown in Appendix B, for given sample size n, the amount of threshold overestimation increases with decreasing shape parameter ξ. The latter determines the convergence rate of the log-transformed GP quantiles to those of an exponential distribution, as the exceedance probability level decreases.

To illustrate the above arguments, Figure 4a shows a Pareto quantile plot obtained from a synthetic sample of size 10^5, drawn from a GP distribution with threshold $u=0$, scale parameter $a_0:=a_{u=0}=10$, and shape parameter $\xi=0.1$. As discussed in section 3.3 below, values of the shape parameter ξ on the order of 0.1–0.2 are representative for rainfall applications [see also *Koutsoyiannis*, 2004b; *Deidda and Puliga*, 2006; *Deidda*, 2007, 2010; *Papalexiou and Koutsoyiannis*, 2013; *Serinaldi and Kilsby*, 2014]. One clearly sees that even in the case of 10^5 sample values (i.e., corresponding to more than 1350 years of daily rainfall data, with an average wet-day fraction of 20%), the threshold u_L^* estimated using the modified Lewis kernel statistic in equation (14) at the 5% significance level is much larger than zero (i.e., $u_L^*=56.9$; for the modified Jackson kernel statistic—not shown here—the estimated threshold is: $u_J^*=58.9$), justifying that linearity in Pareto quantile plots is a much stricter condition than the applicability of a GP distribution model, leading to significant overestimation of the corresponding thresholds.

An additional observation one makes is that even for large threshold values (i.e., $u_L^*=56.9$) obtained using the modified Lewis kernel statistic, Hill's estimator in equation (13) overestimates the shape parameter of the GP distribution by 120% (i.e., $\xi_L=0.221$ instead of 0.1). This becomes apparent if one compares the dashed black line in Figure 4a, corresponding to the slope calculated by applying Hill's estimator in equation (13) to the excesses above threshold $u_L^*=56.9$, to the black solid line, which refers to the theoretical slope. Similar findings have been obtained, also, when using the modified Jackson kernel statistic (not

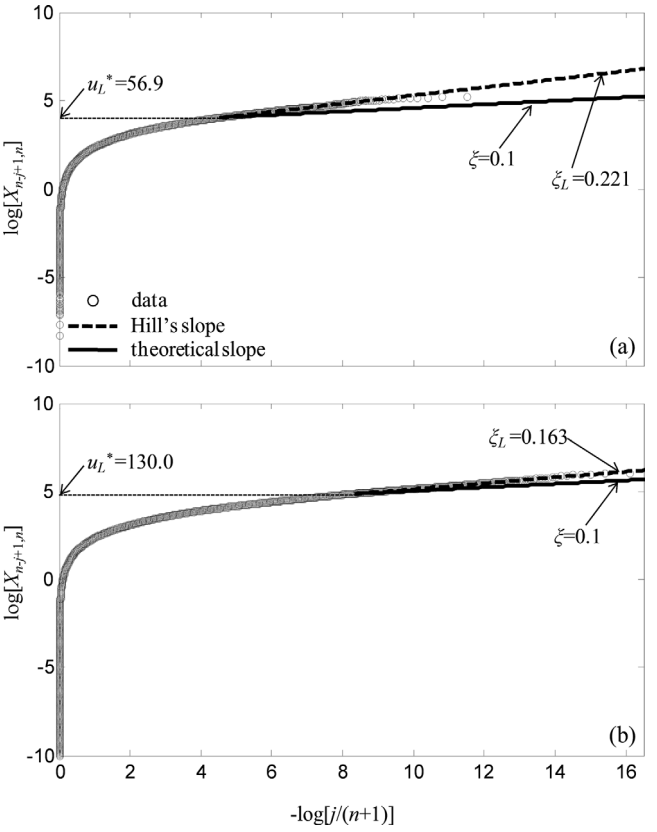

Figure 4. (a) Application of Pareto quantile plot to a synthetic sample of size $n = 10^5$, drawn from a GP distribution with threshold $u = 0$, scale parameter $a_0 := a_{u=0} = 10$, and shape parameter $\xi = 0.1$. The black line denotes the theoretical slope (i.e., $\xi = 0.1$), whereas the dashed line corresponds to that obtained by applying Hill's estimator (equation (13)) above threshold u_L^*. (b) Same as Figure 4a but for sample size $n = 10^7$.

shown here for brevity), where Hill's estimator applied to the excesses above threshold $u_J^* = 58.9$ leads to $\xi_J = 0.226$.

Figure 4b shows similar results to Figure 4a, but for the case of a synthetic sample of size 10^7 (i.e., more than 135,000 years of daily rainfall data, with an average wet-day fraction of 20%). The detected thresholds are considerably larger this time (i.e., $u_L^* = 130.0$ and $u_J^* = 127.1$), indicating the approximate validity of the obtained estimates under preasymptotic conditions, as the calculated thresholds increase considerably with increasing sample size. This is further supported by the fact that the relative error of the obtained Hill estimates remains very high, on the order of 60% (i.e., $\xi_L = 0.163$ and $\xi_J = 0.165$, instead of 0.1), indicating that even 135,000 years of daily rainfall observations do not suffice for the asymptotic conditions to be reached.

Evidently, for low shape parameter values ξ (i.e., on the order of 0.1–0.2), commonly met in rainfall applications, the slow convergence of the log-transformed GP quantiles to those of an exponential distribution with increasing probability level leads to significant overestimation of the calculated thresholds and GP shape parameters. For the interested reader, a detailed theoretical investigation of the above remarks is given in Appendix B.

3. Application of GP Threshold Detection Methods to NOAA-NCDC Rainfall Data Set

3.1. Data

To apply the threshold detection methods reviewed in section 2, we use daily rainfall data from the NOAA-NCDC (National Oceanic and Atmospheric Administration-National Climatic Data Center) open-access database, http://www.ncdc.noaa.gov/oa/climate/ghcn-daily. The latter includes daily rainfall records collected from more than 75,000 stations in 180 countries and territories, located (mainly) in the United States, South Pacific, and Europe. Both the record length and period of record vary by station and cover intervals ranging from less than a year to more than 175 years. A complete list of the stations, together with their geographical coordinates and other metadata, can be found at ftp://ftp.ncdc.noaa.gov/pub/data/ghcn/daily/.

To ensure validity of the obtained results, in our analysis, we considered only 1714 stations with more than 110 years of available observations, and then tested the robustness of each threshold detection method by randomly selecting and eliminating daily values corresponding to 70 years of observations.

Since data quantization affects some of the obtained results (see section 2.3), we used the rounding-off rule estimator (RRE) developed by *Deidda* [2007] to estimate the level of quantization in the rainfall records analyzed. We found that more than 80% of the analyzed stations exhibit quantization levels on the order of $\Delta = 0.1$ mm/d, for a fraction of the recordings that in most cases exceeds 75%. The remaining data fractions

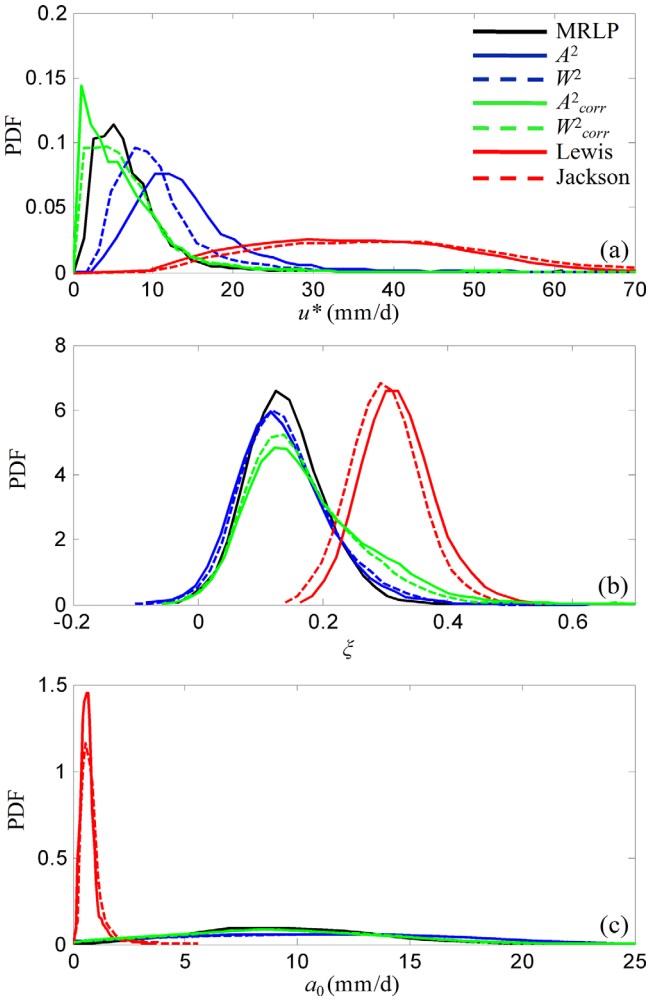

Figure 5. (a) Probability density function of thresholds u^*, estimated by applying the reviewed methods to 1714 rainfall records from NOAA-NCDC database, with more than 110 years of available observations; see main text for details. (b) Same as Figure 5a but for the shape parameter ξ. (c) Same as Figure 5b but for the reparameterized scale parameter a_0.

were found to exhibit quantization levels between 0.1 and 0.2 mm/d as, in all analyzed records, quantization levels larger than $\Delta = 0.2$ mm/d were below the detection level. These findings indicate that the analyzed data are not heavily quantized, but as shown next, even in this case, quantization affects some of the obtained results.

3.2. Methodology

As noted in section 2.1, the nonparametric method of *Gerstengarbe and Werner* [1989] is theoretically incorrect and may lead to strange results, especially in the case of quantized samples. Hence, in what follows, we focus solely on applying the remaining three methods.

For the mean residual life plot (MRLP) method, we followed the procedure described in section 2.2, where the optimal threshold u^* to extract PDS from data is determined as the lowest threshold that corresponds to a local minimum of the weighted mean square error (WMSE) function of the associated linear regression. While the above rule cannot be considered unique, as it is not based on statistical arguments (i.e., a limitation common to all graphical methods), it produces acceptable results for all stations analyzed, allowing for direct comparisons with the remaining threshold detection methods.

When applying the "failure-to-reject" method of *Choulakian and Stephens* [2000, 2001] (see section 2.3) to each rainfall record, we simultaneously increased the threshold used to extract the PDS (i.e., starting from the smallest nonzero sample value), until the null hypothesis of GP distributed excesses was not rejected at

Table 3. Ensemble Mean Value and Standard Deviation of GP Model Parameters, Estimated by Applying the Reviewed Methods to 1714 Rainfall Records From NOAA-NCDC Database, With More Than 110 Years of Available Observations

		Methods						
		MRLP [1]	A^2 [2]	W^2 [3]	A^2_{corr} [4]	W^2_{corr} [5]	Lewis [6]	Jackson [7]
Threshold u^* (mm/d)	Mean	6.68	13.91	10.11	6.50	6.37	34.68	39.43
	St. dev.	6.69	8.14	7.25	7.47	6.59	16.71	20.01
Shape ξ	Mean	0.140	0.138	0.144	0.177	0.167	0.320	0.301
	St. dev.	0.060	0.088	0.096	0.096	0.092	0.054	0.053
Scale a_u (mm/d)	Mean	10.12	10.94	10.35	9.22	9.38	11.54	12.30
	St. dev.	5.03	4.37	4.27	4.54	4.41	5.62	6.21
a_0 (mm/d)	Mean	9.17	8.96	8.77	8.25	8.45	0.67	0.75
	St. dev.	4.36	5.43	6.08	4.28	4.14	0.40	0.59

the 5% significance level. Fitting of the GP distribution model to rainfall excesses above different thresholds was done using the maximum likelihood (ML) method, as suggested by *Grimshaw* [1993].

To study the effects of data quantization on the estimated threshold values, we first applied the "failure-to-reject" method using the W^2 and A^2 quantiles reported in *Choulakian and Stephens* [2001, Table 2] for continuous data, and then using the corresponding quantiles referring to the level of data quantization present in the historical series. The latter were obtained independently for each of the analyzed records, following the Monte Carlo approach described in section 2.3.

Concerning Hill-based methods (i.e., the modified Jackson and Lewis kernel statistics; see section 2.4), the optimal threshold to extract PDS from data was estimated by simultaneously increasing the threshold (i.e., starting from the smallest nonzero sample value), until the null hypothesis of exponentially distributed excesses of the log-transformed PDS (i.e., suggesting linearity of the Pareto quantile plot) was not rejected at the 5% significance level.

For the MRLP and "failure-to-reject" methods, estimation of the shape ξ and scale a_u parameters above different thresholds u was done using the ML method, as suggested by *Grimshaw* [1993] for (a_u, ξ) unknown. For the modified Jackson and Lewis kernel statistics, estimation of the GP shape parameter was done using Hill's estimator in equation (13), while estimation of the scale parameter was done using the *Grimshaw* [1993] ML estimator for known ξ.

3.3. Results and Discussion

Table 3 and Figure 5 summarize the results obtained when applying the threshold detection methods reviewed in sections 2.2–2.4 to the 1714 rainfall records with more than 110 years of empirical observations from the NOAA-NCDC data set.

Table 3 shows the ensemble mean value and standard deviation of the estimated thresholds u^*, together with the ensemble mean values and standard deviations of the shape ξ and scale a_{u^*} parameters of a GP distribution model fitted to rainfall excesses above u^*. Additional insertions that allow for direct interpretation of the obtained results are the ensemble mean value and standard deviation of the reparameterized GP scale parameter $a_0 = a_{u^*} - \xi u^*$ to zero threshold (as suggested by *Deidda* [2010]; see also equation (4) and Appendices A and B). Since the scale parameter a_u of a GP distribution model is linked to the threshold u through equation (4), the aforementioned reparameterization allows for direct comparisons of the results obtained from methods producing significantly different threshold estimates (e.g., the Jackson and Lewis modified kernel statistics relative to the remaining methods).

An important observation one makes is that the estimated thresholds using the Jackson and Lewis modified kernel statistics (see last two columns in Table 3) are more than 3 times larger relative to those calculated by the remaining methods (see columns [1]–[5] in Table 3), while the associated Hill's shape parameter estimates are 2 times larger than those produced by other methods. As noted in section 2.4 (see also Appendix B for a theoretical investigation), the observed differences are due to the slow convergence of a logarithmically transformed GP variable to an exponential distribution, when the GP shape parameter ξ is small; i.e., on the order of 0.1–0.2, commonly met in rainfall applications (see columns [1]–[5] in Table 3 and *Koutsoyiannis* [2004b], *Deidda and Puliga* [2006], *Deidda* [2007, 2010], *Papalexiou and Koutsoyiannis* [2013], and *Serinaldi and Kilsby* [2014]). To visually illustrate the aforementioned results, Figures 5a–5c present the probability density functions (PDFs) of thresholds u^*, shape parameters ξ, and reparameterized scale parameters a_0 resulting from application of each of the reviewed methods to the 1714 analyzed series. One sees the considerable shifts of all distributions of GP model parameters, when estimation is done based on GP asymptotic arguments (i.e., Hill's assumption; see section 2.4 and Appendix B).

Concerning the MRLP graphical method, and the Crámer-von Mises (W^2) and Anderson-Darling (A^2) GoF methods (i.e., applied using the asymptotic distributions for continuous data [*Choulakian and Stephens*, 2001, Table 2]), they all produce results that are of the same order; see columns [1]–[3] in Table 3, Figures 5a–5c, and discussion below. This is due to the fact that the theoretical bases of the three methods lie on properties of the GP distribution model valid, also, under preasymptotic conditions. Two things to notice are: (a) the Anderson-Darling (A^2) GoF method produces slightly higher thresholds relative to Crámer-von Mises (W^2); compare columns [2] and [3] in Table 3 and the solid and dashed blue lines in Figures 5a.

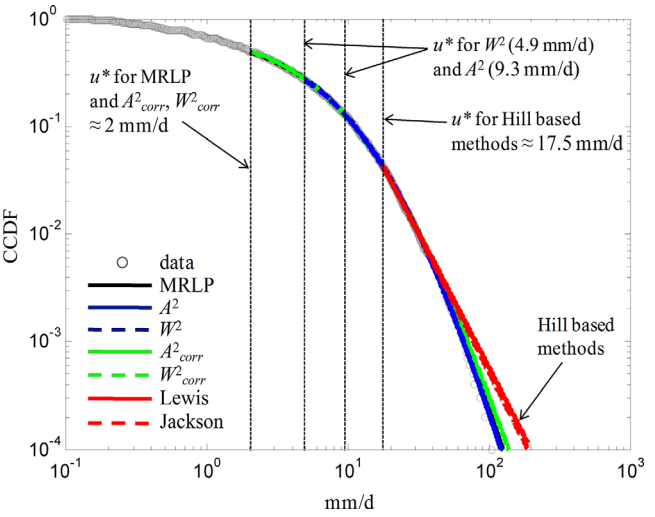

Figure 6. Log-log plots of the empirical and theoretical complementary cumulative distribution functions (CCDFs) obtained for the positive rainrates of the 126 year daily rainfall record from Australia, used also in Figures 1b, 2, and 3.

(b) Both Anderson-Darling (A^2) and Cámer-von Mises (W^2) GoF methods produce slightly higher thresholds relative to the MRLP graphical method (i.e., column [1] in Table 3 and solid black line in Figure 5a).

The first finding should be attributed to the fact that the Anderson-Darling A^2 statistic assigns larger weights to observations located at the (upper or lower) distribution tail [*Choulakian and Stephens*, 2001] and, hence, leads to more conservative (i.e., higher) threshold estimates. The second finding is related to the quantization present in the historical rainfall series. To illustrate this, we reapplied the Anderson-Darling and Crámer-von Mises GoF methods to all rainfall records, using numerically derived distributions (i.e., based on Monte Carlo simulations; see sections 3.2 and 2.3) accounting for the level of quantization present in each of the analyzed series. We refer to these results in Table 3 (i.e., columns [4] and [5]) and Figures 5a–5c (solid and dashed green lines) as A^2_{corr} and W^2_{corr}, respectively. One sees that when the effect of data quantization is incorporated in the corresponding statistics used for threshold detection, the differences in the calculated thresholds using the MRLP and the Crámer-von Mises and Anderson-Darling GoF methods become notably smaller. However, the aforementioned reduction of the calculated thresholds when accounting for data quantization is accompanied by a slight increase (decrease) of the estimated GP shape (scale) parameter. We have investigated this issue in some detail, and found that it is related to a slight increase of the frequency of small threshold estimates (i.e., below 1–1.5 mm/d), which causes the ensemble mean value of the estimated shape parameters to increase.

To illustrate the quality of GP distribution fits to empirical data using the reviewed methods, Figure 6 shows log-log plots of the empirical and theoretical CCDFs obtained for the positive rainrates of the 126 year daily rainfall record from Australia, used also in Figures 1b, 2, and 3. One sees that all methods produce similarly good fits, except for the Jackson and Lewis modified kernel statistics. The latter two are based on Hill's assumption, and lead to much higher GP threshold and shape parameter estimates, causing significant overestimation of the empirical quantiles.

To study the robustness of the reviewed methods on the number of available years in record, for each of the analyzed series, we randomly eliminated daily values corresponding to 70 years of observations, and repeated the aforementioned analysis. The obtained results are summarized in Table 4 and Figure 7.

The first thing to notice is that the results of MRLP are almost identical before and after random thinning of the empirical records, indicating the robustness of the method in record length variations. Concerning the modified Jackson and Lewis kernel statistics (i.e., compare the results in the last two columns of Tables 3 and 4, and the red solid and dashed lines in Figures 5 and 7), reduced sample sizes result in smaller thresholds and larger shape parameter estimates. This is in accordance with the theoretical remarks made in section 2.4, as for larger samples one moves farther into the upper tail of the distribution, with subsequent reduction of the estimated slope (i.e., GP shape parameter) from the Pareto quantile plot (see Figures 4a and 4b and discussion in section 2.4), whereas for small sample sizes, the validity of asymptotic assumptions weakens, making the obtained results less accurate (see Appendix B).

Regarding the "failure-to-reject" method of *Choulakian and Stephens* [2000, 2001], one sees a decrease of the ensemble mean value of the calculated thresholds, with decreasing sample size (i.e., compare the results in columns [2]–[5] of Tables 3 and 4, as well as the blue and green solid and dashed lines in

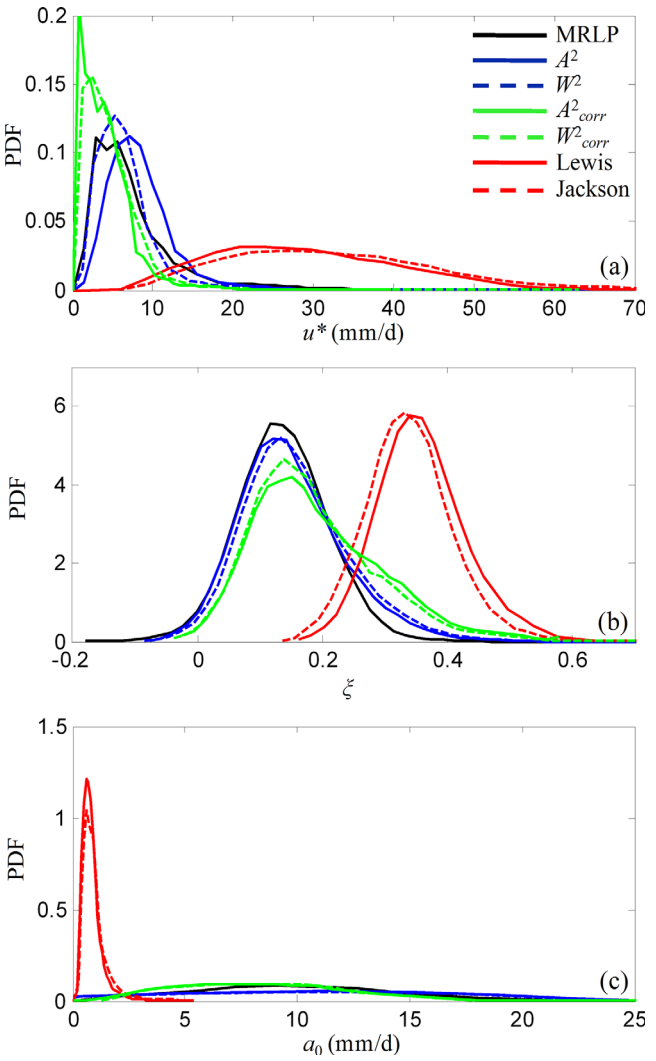

Figure 7. Same as Figure 5, but after randomly eliminating daily values corresponding to 70 years of observations from all analyzed series.

Figures 5 and 7). This is true independent of whether one accounts for quantization in the empirical samples (i.e., columns [4] and [5] in Table 4, and green solid and dashed lines in Figure 7) or not (i.e., columns [2] and [3] in Table 4, and blue solid and dashed lines in Figure 7). The decrease in the obtained threshold estimates is related to the higher probability of type 2 errors of the "failure-to-reject" method with decreasing sample size [see *Choulakian and Stephens*, 2001, section 4 and Table 6], which leads to less conservative (i.e., smaller) threshold estimates.

In summary, the obtained results show that for rainfall applications, where the GP shape parameter ξ is on the order of 0.1–0.2 (see Tables 3 and 4), Hill-assumption-based kernel statistics lead to significantly higher GP threshold and shape parameter estimates. The "failure-to-reject" method of *Choulakian and Stephens* [2000, 2001], which is based on the goodness of fit of a GP distribution to rainfall excesses above different thresholds, produces similar results to the MRLP method, but it is quite sensitive to the presence of quantization in the empirical data, while its statistical power is reduced as the sample size decreases. Finally, the MRLP graphical method is less sensitive to variations of empirical record lengths, and to low levels of data quantization; i.e., for values of Δ/a_0 on the order of 0.01–0.02 used in this study.

Overall, in the light of the theoretical bases and limitations of each of the reviewed methods, and the results presented in Tables 3 and 4 and Figures 5 and 7, one concludes that: (a) the existence of quantization in rainfall records, along with variations in their length, constitute the two most important

Table 4. Same as Table 3, but After Randomly Eliminating Daily Values Corresponding to 70 Years of Observations From All Analyzed Series

		Methods						
		MRLP [1]	A^2 [2]	W^2 [3]	A^2_{corr} [4]	W^2_{corr} [5]	Lewis [6]	Jackson [7]
Threshold u^*(mm/d)	Mean	6.63	7.53	5.53	3.67	3.69	27.96	31.33
	St. dev.	6.70	4.22	3.74	3.29	3.32	13.49	15.00
Shape ξ	Mean	0.137	0.150	0.159	0.192	0.186	0.359	0.338
	St. dev.	0.073	0.096	0.094	0.103	0.103	0.068	0.066
Scale a_u (mm/d)	Mean	10.14	9.78	9.31	8.50	8.61	10.51	11.15
	St. dev.	5.12	3.64	3.65	3.94	3.86	4.95	5.37
a_0 (mm/d)	Mean	9.29	8.62	8.42	7.94	8.04	0.79	0.87
	St. dev.	4.78	4.88	4.60	3.69	3.61	0.50	0.63

factors that (depending on the method used) may significantly affect the accuracy of the obtained results. (b) For daily rainfall applications, GP threshold estimates range between 2 and 12 mm/d, with a mean value around 6.5 mm/d, leading to GP shape ξ and reparameterized scale a_0 parameter estimates on the order of 0.15 and 9 mm/d, respectively. The observed variation of threshold estimates may be partially attributed to the dependence of rainfall statistics on large-scale climatic features (e.g., near-surface temperature [see *Wasko et al*., 2015; *Wasko and Sharma*, 2015]) observed at different locations of the globe. (c) While several studies [see e.g., *Madsen et al*., 1997a,b; *Acero et al*., 2011; *Villarini et al*., 2011, 2013; *Serinaldi and Kilsby*, 2014, among others] have used high enough thresholds (i.e., maintaining less than 5% of the empirical observations) to extract PDS from data, much lower threshold values are also effective leading to reduced estimation variance of GP distribution parameters. For the 1714 overcentennial daily rainfall records analyzed in this study, the mean annual exceedance rate estimated using the MRLP method is on the order of 40 year^{-1}, corresponding to the upper 11% of the empirical observations.

4. Conclusions

To quantify hydrological risks, one needs to accurately model extreme rainfall. Among different approaches that have been suggested, the use of partial duration series (PDS) with a generalized Pareto (GP) distribution model fitted to rainfall excesses above a properly selected threshold u^* has attracted much attention in recent years [see e.g., *Wang*, 1991; *Rosbjerg and Madsen*, 1992, 1995; *Madsen et al*., 1997a, 1997b, 2002; *Lang et al*., 1999; *Willems*, 2000; *Deidda and Puliga*, 2006; *Cooley et al*., 2007; *Veneziano et al*., 2007; *Deidda*, 2010; *Serinaldi and Kilsby*, 2014]. This can be attributed to the fact that it is supported by both theoretical arguments [see e.g., *Balkema and de Haan*, 1974; *Pickands*, 1975; *Leadbetter et al*., 1983; *Smith*, 1985; *Leadbetter*, 1991; *Stedinger et al*., 1993; *Coles*, 2001; *Lucarini et al*., 2016] and empirical evidences [see e.g., *Martins and Stedinger*, 2001a, 2001b; *Deidda and Puliga*, 2006; *Veneziano et al*., 2007, 2009; *Deidda*, 2010; *Veneziano and Yoon*, 2013; *Papalexiou et al*., 2013; *Serinaldi and Kilsby*, 2014], while retaining a larger portion of the empirical data relative to extracting and analyzing solely annual maxima; see section 1.

To consistently apply the PDS/GP approach for extreme excess (EE) modeling, one needs to determine a proper threshold u^* to extract PDS from data, and fit a GP distribution model to rainfall excesses. Selection of u^* is a nontrivial issue, as the selected threshold should be high enough to ensure validity of the GP assumption and, at the same time, maintain a sufficient amount of data to avoid unnecessary increase of the estimation variance of GP distribution parameters.

In this study, we reviewed the theoretical bases of representative GP threshold detection methods commonly met in the relevant literature, discussed their limitations, and applied them to the NOAA-NCDC daily rainfall database.

The nonparametric method suggested by *Gerstengarbe and Werner* [1989] (see also *Werner and Gerstengarbe* [1997], *Domonkos and Piotrowicz* [1998], *Lasch et al*., [1999], *Cebrià n et al*., [2003], and *Cebrià n and Abaurrea* [2006] for representative applications of the method), which uses the sequential version of Mann-Kendal test to detect the starting point of the extreme region of a sample, was proved theoretically incorrect and capable of leading to strange results, especially in the case of quantized samples; see section 2.1.

Hill-assumption-based methods (see section 2.4), although theoretically correct, make use of asymptotic properties of the GP distribution model, which do not always apply under preasymptotic conditions. An example is the case of rainfall, where the low values of the shape parameter ξ (i.e., on the order of 0.1–0.2) cause slow convergence of the log-transformed GP quantiles to those of an exponential distribution. As a consequence, goodness-of-fit methods based on Hill's assumption, such as the Jackson and Lewis modified kernel statistics reviewed in section 2.4, when applied to empirical rainfall records (i.e., of finite length) lead to significant overestimation of GP thresholds and shape parameters.

The "failure-to-reject" method of *Choulakian and Stephens* [2000, 2001], which is based on the recursive application of Crámer-von Mises (W^2) or Anderson-Darling (A^2) goodness-of-fit tests to rainfall excesses above different thresholds, was proved quite sensitive to the presence of quantization in the empirical data, with statistical power that reduces with decreasing sample size; see sections 2.3 and 3.3. The sensitivity of the method even to small levels of data quantization does not allow for routine applications, requiring:

(a) identification of the level of data quantization in the analyzed series (e.g., using the approach of *Deidda* [2007]), and (b) use of Monte Carlo simulations to obtain the required distributions of the Crámer-von Mises (W^2_{corr}) and Anderson-Darling (A^2_{corr}) statistics, as a function of the sample size in the presence of data quantization. Also, for heavily quantized data, no limiting distributions for W^2 and A^2 should exist.

Finally, the mean residual life plot (MRLP) graphical method is quite simple to apply, it is based on GP distribution properties valid (also) under preasymptotic conditions, while demonstrating reduced sensitivity to the length of the available data, and to low levels of data quantization (i.e., for values of Δ/a_0 on the order of 0.01–0.02). The latter should be attributed to the fact that data quantization, unless pronounced, does not alter considerably the mean value of the excesses above different thresholds, making the power of the method rather insensitive to low quantization levels.

Note, however, that the aforementioned competitive advantages of MRLP, relative to the other methods reviewed, were revealed only after the method was brought in a form that avoids visual inspection of the data, facilitating automation. Evidently, there is still room for significant improvements, as the stepwise diagnostic procedure described in section 2.2 to identify the optimal threshold, can be refined to include statistical arguments. This, as well as the maximum level of data quantization that MRLP can be effectively applied, will form the subjects of a future communication.

We conclude by saying that in the case of highly quantized data, estimation of GP model parameters exhibits important convergence-related issues [see e.g., *Deidda and Puliga*, 2009], pointing toward the use of more robust estimation methods such as the Multiple Threshold Method (MTM) suggested by *Deidda* [2010]. MTM fits a GP distribution model to data using a range of possible thresholds, rather than a single one. For the latter purpose, a reliable range of possible threshold estimates for daily rainfall series is between 2 and 12 mm/d, as suggested by the conducted analysis using more than 1700 overcentennial daily rainfall records collected worldwide.

Appendix A: Derivations on the Generalized Pareto (GP) Distribution Model

Define $Z = X_{u^*} = [X - u^* | X > u^*]$ to be the scaled excesses of random variable X above threshold u^*. If Z has generalized Pareto (GP) distribution with complementary cumulative distribution function (CCDF)

$$P[Z > z] = P[Z > z | Z > 0] = \left(1 + \xi \frac{z}{a_{u^*}}\right)^{-1/\xi}, z \geq 0 \tag{A1}$$

where a_{u^*} and ξ are the scale and shape parameters of the distribution, respectively, then for any threshold $u \geq u^*$ and $z \geq u - u^* = u^+ \geq 0$, one has

$$P\left[Z > z | Z > u^+\right] = \frac{P[Z > z \cap Z > u^+]}{P[Z > u^+]} = \frac{P[Z > z]}{P[Z > u^+]} = \frac{P[Z > z | Z > 0]}{P[Z > u^+ | Z > 0]} \tag{A2}$$

By combining equations (A1) and (A2), one obtains

$$P\left[Z > z | Z > u^+\right] = \frac{\left(1 + \xi \frac{z}{a_{u^*}}\right)^{-1/\xi}}{\left(1 + \xi \frac{u^+}{a_{u^*}}\right)^{-1/\xi}} = \left(1 + \frac{\xi(z - u^+)}{a_{u^*} + \xi u^+}\right)^{-1/\xi} = \left(1 + \frac{\xi(z - u^+)}{a_{u^*} + \xi(u - u^*)}\right)^{-1/\xi} \tag{A3}$$

and by defining $Y = Z - u^+ = X - u$, equation (A3) receives the form:

$$P[Y > y] = P[Y > y | Y > 0] = \left(1 + \xi \frac{y}{a_u}\right)^{-1/\xi}, y \geq 0 \tag{A4}$$

where $a_u = a_{u^*} + \xi(u - u^*)$.

Equation (A4) suggests that if the scaled excesses $X_{u^*} = [X - u^* | X > u^*]$ of random variable X above threshold u^* are GP distributed with shape and scale parameters ξ and a_{u^*}, respectively, then the scaled excesses $X_u = [X - u | X > u]$ of random variable X above any threshold $u \geq u^*$ are also GP distributed with the same shape parameter, and scale parameter that depends linearly on u [see also *Coles*, 2001; *Deidda*, 2010]:

$$a_u = a_{u^*} + \xi(u - u^*) \tag{A5}$$

Further, the expected value of the scaled excesses $X_u = [X - u | X > u]$ is calculated to be

$$e(u) = E[X - u \mid X > u] = \int_0^\infty y\, dQ_u(y) = \frac{a_u}{1-\xi} = \frac{a_{u^*} + \xi(u - u^*)}{1-\xi} = Au + B \tag{A6}$$

where $Q_u(y)$ is the CDF in equation (2), a_u and a_{u^*} are the scale parameters of the distribution when fitted to the excesses above thresholds u and u^*, respectively, and $A = \xi/(1-\xi)$ and $B = (a_{u^*} - \xi u^*)/(1-\xi)$ are the slope and intercept of the linear relation.

Appendix B: Hill's Assumption

In section 2.4, we referred to the asymptotic exponentiality of log-transformed Pareto-type variables; i.e., satisfying equations (10–12). Then we linked it to Pareto quantile plots, as a means of detecting the sample value (i.e., threshold) above which the complementary CDF (CCDF) of the data displays a log-log linear behavior (also referred to as Hill's assumption, see section 2.4). Finally, we used Monte Carlo simulations to illustrate the significant biases of the obtained estimates, when threshold detection and shape parameter estimation are done by applying goodness-of-fit (GoF) kernel statistics based on asymptotic arguments to finite samples (i.e., under preasymptotic conditions). In what follows, we use theoretical arguments to introduce an indicator for the threshold and shape parameter overestimation induced by Hill's assumption, in the case of a generalized Pareto (GP) distribution.

Suppose that the scaled excesses $X_u = [X - u | X > u]$ of random variable X above threshold u follow a GP distribution model with shape parameter $\xi > 0$, and scale parameter a_u (see equation (2)):

$$P[X - u \le y | X > u] = 1 - \left(1 + \xi \frac{y}{a_u}\right)^{-1/\xi}, y \ge 0 \tag{B1}$$

Letting $x = y + u$ in equation (B1) and solving for x, one obtains:

$$x_{u,p'} = \frac{a_u}{\xi}\left[(1-p')^{-\xi} - 1\right] + u, x_{u,p'} \ge u \tag{B2}$$

where $x_{u,p'}$ denotes the p'-quantile of random variable $[X|X > u]$.

Based on the GP threshold invariance property discussed in Appendix A, and using equations (B2) and (A5), one can easily show that for any $x_{u,\,p'} \ge u$:

$$x_{u,p'} = t_p = \frac{a_0}{\xi}\left[(1-p)^{-\xi} - 1\right] \tag{B3}$$

where t_p denotes the p-quantile of a GP distributed random variable, T, with shape parameter ξ, zero threshold, and scale parameter $a_0 = a_u - \xi u$. In this case, the probabilities $p'(x) = P[X \le x | X > u]$ and $p(t) = P[T \le t]$ in equations (B2) and (B3), respectively, are linked through:

$$p'(x) = \frac{p(x) - p(u)}{1 - p(u)}, \text{ for any } x \ge u \tag{B4}$$

where $p(t) = P[T \le t] = 1 - \left(1 + \xi \frac{t}{a}\right)^{-1/\xi}$.

Consider now the standardized random variable $T' := \xi T/a_0$, which is also GP distributed with shape and scale parameters equal to ξ. It follows from equation (B3) that the log-transformed p-quantile of T' satisfies:

$$\log(t'_p) = z_p + \log[1 - (1-p)^{\xi}] \tag{B5}$$

where $z_p = -\xi \log(1-p)$ is the p-quantile of an exponentially distributed random variable Z with mean value equal to ξ. The second term at the right-hand side of equation (B5) can be viewed as an additive bias that vanishes as $p \to 1$ or, equivalently, $t'_p \to \infty$. Hence, at the limit as $p \to 1$, the standardized log-transformed GP variable T' is attracted to an exponential distribution, with mean value equal to the GP shape parameter ξ. This corresponds to a linear Pareto quantile plot with slope equal to ξ (see also section 2.4).

Using equation (B5), equation (B3) can be written as:

$$x_{u,p'} = t_p = \frac{a_0}{\xi}\exp\left(l_p z_p\right) \tag{B6}$$

where l_p denotes the quantile ratio:

$$l_p = \log\left(t'_p\right)/z_p = 1 + \frac{\log\left[1-(1-p)^{\xi}\right]}{\log\left[(1-p)^{-\xi}\right]} \tag{B7}$$

An alternative way to view l_p in equation (B7) is as the multiplicative bias of the log-transformed p-quantile of the standardized GP variable T' under preasymptotic conditions, relatively to what one expects asymptotically as $p \to 1$ (i.e., note that as $p \to 1$, $l_p \to 1$). Hence, l_p can be used as an indicator for the level of convergence to GP asymptotics (i.e., linearity in a Pareto quantile plot), under preasymptotic conditions.

It follows from equations (B6) and (B7) that the estimation accuracy of both the threshold u^* and shape parameter ξ from finite samples (i.e., under preasymptotic conditions), by applying the Jackson and Lewis modified kernel statistics and Hill's estimator in equation (13), respectively, is determined by the value of the ratio l_p; where p can be set to the largest empirical probability value in a sample of size n. For rainfall applications, and in the case when a Weibull plotting position formula is used, $p = n_w/(n_w + 1)$, where n_w is the number of positive values in the record (i.e., corresponding to wet conditions).

To study the sensitivity of the ratio l_p in equation (B7) to the GP shape parameter ξ and the size of the sample, Figure B1 shows plots of l_p as a function of the wet-sample size $n_w = p/(1-p)$, for indicative values of ξ. The thick curve has been obtained using $\xi = 0.15$, corresponding to an average value for the GP shape parameter estimates obtained from the NOAA-NCDC data set (see Table 3). One sees that l_p increases with increasing wet-sample size n_w, approaching 1 (i.e., asymptotic conditions) as $n_w \to \infty$. An important note to make, is that the convergence rate of l_p to 1 increases with increasing shape parameter ξ.

For example, in the theoretical case when $\xi \geq 0.3$, convergence to GP asymptotics is quite fast, with approximate validity even in the case of relatively small samples (i.e., $l_p > 0.95$ for $n_w = 2 \times 10^3$; see Figure B1). Under this setting, use of Jackson and Lewis modified kernel statistics and Hill's estimator would lead to relatively accurate GP threshold and shape parameter estimates, even for sample lengths on the order of 30 years (i.e., assuming an average wet-day fraction of 20%). However, for values of $\xi = 0.1$–0.2 descriptive of daily rainfall records (see sections 3 and 4), l_p converges slowly to 1 with increasing wet-sample size n_w, leading to significant biases in the obtained estimates. Since values of ξ on the order of 0.3 and higher are not common in rainfall records, Hill-assumption-based methods (as the Jackson and Lewis modified kernel statistics reviewed here) should be avoided as they are expected to produce significantly biased results.

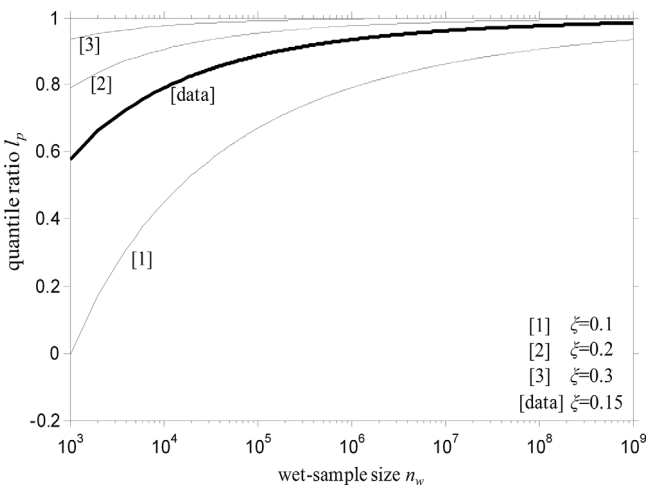

Figure B1. Quantile ratio l_p in equation (B7) as a function of the number of the positive values in record, n_w, for different values of ξ.

Similar concepts to those used in deriving equations (B6) and (B7) can be applied to any Pareto-type distribution, other than the GP. The derivations are beyond the scopes of this paper and are left to the reader.

Acknowledgments
Parts of this work were implemented within the framework of the Action (Supporting Postdoctoral Researchers) of the Operational Program "Education and Lifelong Learning" (Action's Beneficiary: General Secretariat for Research and Technology), and cofinanced by the European Social Fund (ESF) and the Greek State. The work of Andreas Langousis was supported by the Onassis Foundation under the "Special Grant and Support Program for Scholars' Association Members". The work conducted by Roberto Deidda was funded under the Sardinian Regional Law 7/2007 (funding call 2013). The NOAA-NCDC rainfall data set used in this study is freely available at http://www.ncdc.noaa.gov/oa/climate/ghcn-daily.

References

Acero, F. J., J. A. García, and M. C. Gallego (2011), Peaks-over-threshold study of trends in extreme rainfall over Iberian Peninsula, *J. Clim.*, *24*, 1089–1105.
Adamowski, K. (2000), Regional analysis of annual maximum and partial duration flood data by nonparametric and L-moment methods, *J. Hydrol.*, *229*(3-4), 219–231.
Ahmad, M. I., C. D. Sinclair, and B. D. Spurr (1988), Assessment of flood frequency models using empirical distribution function statistics, *Water Resour. Res.*, *24*(8), 1323–1328.
Anderson, T. W., and D. A. Darling (1952), Asymptotic theory of certain goodness-of-fit criteria based on stochastic processes, *Ann. Math. Stat.*, *23*, 193–212.
Anderson, T. W., and D. A. Darling (1954), A test of goodness-of-fit, *J. Am. Stat. Assoc.*, *49*, 765–769.
Balkema, A. A., and L. de Haan (1974), Residual lifetime at great age, *Ann. Probab.*, *2*, 792–804.
Beirlant, J., and J. L. Teugels (1987), Asymptotic normality of Hill's estimator, in *Extreme Value Theory, Oberwolfach 1987*, edited by J. Hüsler and R. D. Reiss, pp. 148–155, Springer, N. Y.
Beirlant, J., P. Vynckier, and J. L. Teugels (1996), Tail index estimation, Pareto quantile plots, and regression diagnostics, *J. Am. Stat. Assoc.*, *91*, 1659–1667.
Beirlant, J., G. Dierckx, Y. Goegebeur, and G. Matthys (1999), Tail index estimation and an exponential regression model, *Extremes*, *2*(2), 177–200.
Beirlant, J., Y. Goegebeur, J. Segers, and J. L. Teugels (2004), *Statistics of Extremes—Theory and Applications, Wiley Ser. Probab. Stat.*, John Wiley, Hoboken, N. J.
Beirlant, J., T. de Wet, and Y. Goegebeur (2006), A goodness-of-fit statistic for Pareto-type behaviour, *J. Comput. Appl. Math.*, *186*(1), 99–116.
Bernard, M. M. (1932), Formulas for rainfall intensities of long durations, *Trans. Am. Soc. Civ. Eng.*, *96*, 592–624.
Cebriàn, A. C., and J. Abaurrea (2006), Drought analysis based on a marked cluster Poisson model, *J. Hydrometeorol.*, *7*, 713–723.
Cebriàn, A. C., M. Denuit, and P. Lambert (2003), Generalized Pareto fit to the Society of Actuaries' large claims database, *North Am. Actuarial J.*, *7*(3), 18–36.
Chavez-Demoulin, V., and A. C. Davison (2012), Modelling time series extremes, *REVSTAT Stat. J.*, *10*(1), 109–133.
Choulakian, V., and M. A. Stephens (2000), Goodness-of-fit tests for the generalized Pareto distribution, research report, Dep. of Math. and Stat., Simon Fraser Univ., Burnaby, B. C., Canada.
Choulakian, V., and M. A. Stephens (2001), Goodness-of-fit tests for the generalized Pareto distribution, *Technometrics*, *43*(4), 478–484.
Coles, S. (2001), *An Introduction to Statistical Modeling of Extreme Values*, Springer, London, U. K.
Coles, S., and A. Tawn (1996), Modeling extremes of the areal rainfall process, *J. R. Stat. Soc., Ser. B*, *58*(2), 329–347.
Coles, S., L. R. Pericchi, and S. Sisson (2003), A fully probabilistic approach to extreme rainfall modeling, *J. Hydrol.*, *273*(1-4), 35–50.
Cooley, D., D. Nychka, and P. Naveau (2007), Bayesian spatial modeling of extreme precipitation return levels, *J. Am. Stat. Assoc.*, *102*(479), 824–840.
Csörgő, S., and D. M. Mason (1985), Central limit theorems for sums of extremes values, *Math. Proc. Cambridge Philos. Soc.*, *98*, 546–558.
Csörgő, S., P. Deheuvels, and D. M. Mason (1985), Kernel estimates of the tail index of a distribution, *Ann. Stat.*, *13*, 1050–1077.
Cunnane, C. (1973), A particular comparison of annual maxima and partial duration series methods of flood frequency prediction, *J. Hydrol.*, *18*(3-4), 257–271.
Das, B., and S. Ghosh (2013), Weak limits for exploratory plots in the analysis of extremes, *Bernoulli*, *19*(1), 308–343.
Davis, R., and S. Resnick (1984), Tail estimates motivated by extreme value theory, *Ann. Stat.*, *12*, 1467–1487.
Davison, A. C., and R. L. Smith (1990), Models for exceedances over high thresholds, *J. R. Stat. Soc., Ser. B*, *52*(3), 393–442.
de Haan, L., and A. Ferreira (2006), *Extreme Value Theory: An Introduction*, Springer, N. Y. [Available at http://link.springer.com/book/10.1007%2F0-387-34471-3.]
Deidda, R. (2000), Rainfall downscaling in a space-time multifractal framework, *Water Resour. Res.*, *36*(7), 1779–1784.
Deidda, R. (2007), An efficient rounding-off rule estimator: Application to daily rainfall time series, *Water Resour. Res.*, *43*, W12405, doi:10.1029/2006WR005409.
Deidda, R. (2010), A multiple threshold method for fitting the generalized Pareto distribution to rainfall time series, *Hydrol. Earth Syst. Sci.*, *14*(12), 2559–2575.
Deidda, R., and M. Puliga (2006), Sensitivity of goodness-of-fit statistics to rainfall data rounding off, *Phys. Chem. Earth*, *31*(18), 1240–1251.
Deidda, R., and M. Puliga (2009), Performances of some parameter estimators of the generalized Pareto distribution over rounded-off samples, *Phys. Chem. Earth*, *34*(10-12), 626–634.
Deidda, R., R. Benzi, and F. Siccardi (1999), Multifractal modeling of anomalous scaling laws in rainfall, *Water Resour. Res.*, *35*(6), 1853–1867.
Deidda, R., M. G. Badas, and E. Piga (2004), Space-time scaling in high intensity Tropical Ocean Global Atmosphere Coupled Ocean-Atmosphere Response Experiment (TOGA-COARE) storms, *Water Resour. Res.*, *40*, W02506, doi:10.1029/2003WR002574.
Deidda, R., M. G. Badas, and E. Piga (2006), Space-time multifractality of remotely sensed rainfall fields, *J. Hydrol.*, *322*, 2–13.
Demarée, G. (1985), Intensity-duration-frequency relationship of point precipitation at Uccle, Reference Period 1934–1983, *Inst. R. Météorol. Belg., Ser. A*, *116*, 1–52.
Domonkos, P., and K. Piotrowicz (1998), Winter temperature characteristics in central Europe, *Int. J. Climatol.*, *18*, 1405–1417.
Dupuis, D. J. (1998), Exceedances over high thresholds: A guide to threshold selection, *Extremes*, *1*(3), 251–261.
Dupuis, D. J., and M. Tsao (1998), A hybrid estimator for generalized Pareto and extreme-value distributions, *Commun. Stat. Theory Methods*, *27*(4), 925–941.
El Adlouni, S., and T. B. M. J. Ouarda (2010), Frequency analysis of extreme rainfall events, in *Rainfall: State of the Science, Geophys. Monogr. Ser.*, vol. 191, 287 pp., AGU, Washington, D. C.
Fisher, R. A., and L. H. C. Tippett (1928), Limiting forms of the frequency distribution of the largest or smallest member of a sample, *Proc. Cambridge Philos. Soc.*, *24*(2), 180–190.
Gerstengarbe, F. W., and P. C. Werner (1989), A method for the statistical definition of extreme-value regions and their application to meteorological time series, *Z. Meteorol.*, *39*(4), 224–226.

Gerstengarbe, F. W., and P. C. Werner (1991), Some critical remarks on the use of extreme-value statistics in climatology, *Theor. Appl. Climatol.*, *44*(1), 1–8.
Gerstengarbe, F. W., and P. C. Werner (1999), Estimation of the beginning and end of recurrent events within a climate regime, *Clim. Res.*, *11*, 97–107.
Gnedenko, B. (1943), Sur la distribution limite du terme maximum d'une serie aleatoire, *Ann. Math.*, *44*(3), 423–453.
Goegebeur, Y., J. Beirlant, and T. de Wet (2008), Linking Pareto-tail kernel goodness-of-fit statistics with tail index at optimal threshold and second order estimation, *REVSTAT*, *6*(1), 51–69.
Grimshaw, S. D. (1993), Computing maximum likelihood estimates for the generalized Pareto distribution, *Technometrics*, *35*, 185–191.
Gumbel, E. J. (1958), *Statistics of Extremes*, Columbia Univ. Press, N. Y.
Haeusler, E., and J. L. Teugels (1985), On asymptotic normality of Hill's estimator for the exponent of regular variation, *Ann. Stat.*, *13*, 743–756.
Hall, P. (1982), On some estimates of an exponent of regular variation, *J. R. Stat. Soc., Ser. B*, *44*, 37–42.
Hall, P. (1990), Using the bootstrap to estimate mean squared error and select smoothing parameter in nonparametric problems, *J. Multivariate Anal.*, *32*, 177–203.
Hall, P., and A. H. Welsh (1985), Adaptive estimates of parameters of regular variation, *Ann. Stat.*, *13*(1), 331–341.
Hampel, F. R. (1968), Contributions to the theory of robust estimation, PhD thesis, Univ. of Calif., Berkeley.
Hampel, F. R. (1974), The influence curve and its role in robust estimation, *J. Am. Stat. Assoc.*, *69*, 383–393.
Hampel, F. R., E. M. Ronchetti, P. J. Rousseeuw, and W. A. Stahel (1986), *Robust Statistics: The Approach Based on Influence Functions*, John Wiley, N. Y.
Hill, B. M. (1975), A simple general approach to inference about the tail of a distribution, *Ann. Stat.*, *3*(5), 1163–1174.
Hosking, J. R. M., and J. R. Wallis (1987), Parameter and quantile estimation for the generalized Pareto distribution, *Technometrics*, *29*, 339–349.
Hubert, P., Y. Tessier, S. Lovejoy, D. Schertzer, F. Schmitt, P. Ladoy, J. P. Carbonnel, S. Violette, and I. Desurosne (1993), Multifractals and extreme rainfall events, *Geophys. Res. Lett.*, *20*(10), 931–934.
Huber, P. J. (1964), Robust estimation of a location parameter, *Ann. Math. Stat.*, *35*, 73–101.
Jackson, O. A. Y. (1967), An analysis of departures from the exponential distribution, *J. R. Stat., Soc. B*, *29*(3), 540–549.
Jenkinson, A. F. (1955), The frequency distribution of the annual maximum (or minimum) values of meteorological elements, *Q. J. R. Meteorol. Soc.*, *81*(348), 158–171.
Juarez, S. F., and W. R. Schucany (2004), Robust and efficient estimator for the generalized Pareto distribution, *Extremes*, *7*, 237–251.
Karpouzos, D. K., S. Kavalieratou, and C. Babajimopoulos (2010), Trend analysis of precipitation data in Pieria region (Greece), *Eur. Water*, *30*, 31–40.
Katz, R., M. Parlange, and P. Naveau (2002), Statistics of extremes in hydrology, *Adv. Water Resour.*, *25*(8-12), 1287–1304.
Kendall, M. G. (1938), A new measure of rank correlation, *Biometrika*, *30*(1/2), 81–93.
Koutsoyiannis, D. (2004a), Statistics of extremes and estimation of extreme rainfall: I. Theoretical investigation, *Hydrol. Sci. J.*, *49*(4), 575–590.
Koutsoyiannis, D. (2004b), Statistics of extremes and estimation of extreme rainfall: II. Empirical investigation of long rainfall records, *Hydrol. Sci. J.*, *49*(4), 591–610.
Koutsoyiannis, D., D. Kozonis, and A. Manetas (1998), A mathematical framework for studying rainfall intensity-duration-frequency relationships, *J. Hydrol.*, *206*, 118–135.
Kratz, M., and S. I. Resnick (1996), The qq-estimator and heavy tails, *Commun. Stat. Stochastic Models*, *12*(4), 699–724.
Laio, F. (2004), Cramer-von Mises and Anderson-Darling goodness of fit tests for extreme value distributions with unknown parameters, *Water Resour. Res.*, *40*, W09308, doi:10.1029/2004WR003204.
Lang, M., T. B. M. J. Ouarda, and B. Bobée (1999), Towards operational guidelines for over-threshold modeling, *J. Hydrol.*, *225*(3-4), 103–117.
Langousis, A., and D. Veneziano (2007), Intensity-duration-frequency curves from scaling representations of rainfall, *Water Resour. Res.*, *43*, W02422, doi:10.1029/2006WR005245.
Langousis, A., D. Veneziano, P. Furcolo, and C. Lepore (2009), Multifractals rainfall extremes: Theoretical analysis and practical estimation, *Chaos Solitons Fractals*, *39*(3), 1182–1194.
Langousis, A., A. A. Carsteanu, and R. Deidda (2013), A simple approximation to multifractal rainfall maxima using a generalized extreme value distribution model, *Stochastic Environ. Res. Risk Asses.*, *27*(6), 1525–1531.
Lasch, P., M. Lindner, B. Ebert, M. Flechsig, F. W. Gerstengarbe, F. Suckow, and P. C. Werner (1999), Regional impact analysis of climate change on natural and managed forests in the Federal State of Brandenburg, Germany, *Environ. Model. Assess.*, *4*(4), 273–286.
Leadbetter, M. R. (1991), On a basis for ''peak over threshold'' modeling, *Stat. Probab. Lett.*, *12*, 357–362.
Leadbetter, M. R., G. Lindgren, and H. Rootzen (1983), *Extremes and Related Properties of Random Sequences and Series*, Springer, N. Y.
Lewis, P. A. W. (1965), Some results on tests for Poisson processes, *Biometrika*, *52*, 67–77.
Lovejoy, S., and D. Schertzer (1995), Multifractals and rain, in *New Uncertainty Concepts in Hydrology and Hydrological Modelling*, edited by A. W. Kundzewicz, pp. 62–103, Cambridge Press, Cambridge, U. K.
Lucarini, V., D. Faranda, A. C. G. M. M. de Freitas, J. M. M. de Freitas, M. Holland, T. Kuna, M. Nicol, M. Todd, and S. Vaienti (2016), *Extremes and Recurrence in Dynamical Systems*, 304 pp., John Wiley, Hoboken, N. J.
Maasch, K. A. (1988), Statistical detection of the mid-Pleistocene transition, *Clim. Dyn.*, *2*(3), 133–143.
Madsen, H., P. F. Rasmussen, and D. Rosbjerg (1997a), Comparison of annual maximum series and partial duration series methods for modeling extreme hydrologic events: 1. At-site modeling, *Water Resour. Res.*, *33*(4), 747–757.
Madsen, H., C. P. Pearson, and D. Rosbjerg (1997b), Comparison of annual maximum series and partial duration series methods for modeling extreme hydrologic events: 2. Regional modeling, *Water Resour. Res.*, *33*(4), 759–769.
Madsen, H., P. S. Mikkelsen, D. Rosbjerg, and P. Harremoes (2002), Regional estimation of rainfall intensity-duration-frequency curves using generalized least squares regression of partial duration series statistics, *Water Resour. Res.*, *38(*11), 1239, doi:10.1029/2001WR001125.
Mann, H. B. (1945), Nonparametric tests against trend, *Econometrica*, *13*(3), 245–259.
Martins, E. S., and J. R. Stedinger (2001a), Historical information in a generalized maximum likelihood framework with partial duration and annual maximum series, *Water Resour. Res.*, *37*(10), 2559–2567.
Martins, E. S., and J. R. Stedinger (2001b), Generalized maximum likelihood Pareto-Poisson estimators for partial duration series, *Water Resour. Res.*, *37*(10), 2551–2557.
Papalexiou, S. M., and D. Koutsoyiannis (2013), Battle of extreme value distributions: A global survey on extreme daily rainfall, *Water Resour. Res.*, *49*, 187–201, doi:10.1029/2012WR012557.

Papalexiou, S. M., D. Koutsoyiannis, and C. Makropoulos (2013), How extreme is extreme? An assessment of daily rainfall distribution tails, *Hydrol. Earth Syst. Sci.*, *17*(2), 851–862.
Pickands, J. (1975), Statistical inference using extreme order statistics, *Ann. Stat.*, *3*, 119–131.
Resnick, S., and C. Stărică (1997), Smoothing the Hill estimator, *Adv. Appl. Probab.*, *29*(1), 271–293.
Rosbjerg, D., and H. Madsen (1992), On the choice of threshold level in partial duration series, in *Nordic Hydrological Conference, Alta*, *NHP Rep. 30*, edited by G. Østrem, pp. 604–615, Coord. Comm. for Hydrol. in the Nordic Countries, Oslo.
Rosbjerg, D., and H. Madsen (1995), Uncertainty measures of regional flood frequency estimators, *J. Hydrol.*, *167*, 209–224.
Rosbjerg, D., and H. Madsen (2004), Advanced approaches in PDS/POT modeling of extreme hydrological events, in *Hydrology: Science and Practice for 21st Century*, edited by Webb et al., vol. 1, pp. 217–220, Proceedings of the British Hydrological Society International Conference, Imperial College, London, U. K.
Scarrott, C., and A. MacDonald (2012), A review of extreme value threshold estimation and uncertainty quantification, *Stat. J.*, *10*(1), 33–60.
Schertzer, D., and S. Lovejoy (1987), Physical modeling and analysis of rain and clouds by anisotropic scaling of multiplicative processes, *J. Geophys. Res.*, *92*(D8), 9693–9714.
Serinaldi, F., and C. G. Kilsby (2014), Rainfall extremes: Toward reconciliation after the battle of distributions, *Water Resour. Res.*, *50*, 336–352, doi:10.1002/2013WR014211.
Smith, R. L. (1985), Maximum likelihood estimation in a class of non-regular cases, *Biometrika*, *72*, 67–90.
Sneyers, R. (1963), Sur la détermination de la stabilité des séries climatologiques, in *Proceedings of UNESCO-WMO Symposium on Changes of Climate, UNESCO Arid Zone Res. Ser.*, vol. 20, pp. 37–40, Unesco, Paris.
Sneyers, R. (1975), Sur l'analyse statistique des séries d'observations, *Tech. Note 143*, WMO, Geneva, Switzerland.
Stedinger, J. R., R. M. Vogel, and E. Foufoula-Georgiou (1993), Frequency analysis of extreme events, in *Handbook of Applied Hydrology* edited by D. A. Maidment, chap. 18, pp. 18-1–18-66, McGraw-Hill, N. Y.
Stephens, M. A. (1986), Tests based on EDF statistics, in *Goodness-Of-Fit Techniques*, edited by D'Agostino, B. Ralph, and M. A. Stephens, chap. 4, pp. 97–193, Marcel Dekker, N. Y.
Tanaka, S., and K. Takara (2002), A study on threshold selection in POT analysis of extreme floods, *IAHS AISH Pub.*, *271*, 299–304.
Tancredi, A., C. Anderson, and A. O'Hagan (2006), Accounting for threshold uncertainty in extreme value estimation, *Extremes*, *9*(2), 87–106, doi:10.1007/s10687-006-0009-8.
Vandewalle, B., J. Beirlant, A. Christmann, and M. Hubert (2007), A robust estimator for the tail index of Pareto-type distributions, *Comput. Stat. Data Anal.*, *51*(12), 6252–6268.
Veneziano, D., and A. Langousis (2005), The areal reduction factor a multifractal analysis, *Water Resour. Res.*, *41*, W07008, doi:10.1029/2004WR003765.
Veneziano, D., and A. Langousis (2010), Scaling and fractals in hydrology, in *Advances in Data-Based Approaches for Hydrologic Modeling and Forecasting*, edited by B.Sivakumar and R. Berndtsson, 145 pp., World Sci. [Available at http://www.worldscientific.com/worldscibooks/10.1142/7783.]
Veneziano, D., and C. Lepore (2012), The scaling of temporal rainfall, *Water Resour. Res.*, *48*, W08516, doi:10.1029/2012WR012105.
Veneziano, D., and S. Yoon (2013), Rainfall extremes, excesses, and intensity-duration-frequency curves: A unified asymptotic framework and new nonasymptotic results based on multifractal measures, *Water Resour. Res.*, *49*, 4320–4334, doi:10.1002/wrcr.20352.
Veneziano, D., A. Langousis, and P. Furcolo (2006), Multifractality and rainfall extremes: A review, *Water Resour. Res.*, *42*, W06D15, doi:10.1029/2005WR004716.
Veneziano, D., C. Lepore, A. Langousis, and P. Furcolo (2007), Marginal methods of intensity-duration-frequency estimation in scaling and nonscaling rainfall, *Water Resour. Res.*, *43*, W10418, doi:10.1029/2007WR006040.
Veneziano, D., A. Langousis, and C. Lepore (2009), New asymptotic and preasymptotic results on rainfall maxima from multifractal theory, *Water Resour. Res.*, *45*, W11421, doi:10.1029/2009WR008257.
Villarini, G., J. A. Smith, A. A. Ntelekos, and U. Schwarz (2011), Annual maximum and peaks-over-threshold analyses of daily rainfall accumulations for Austria, *J. Geophys. Res.*, *116*, D05103, doi:10.1029/2010JD015038.
Villarini, G., J. A. Smith, and G. A. Vecchi (2013), Changing frequency of heavy rainfall over the central United States, *J. Clim.*, *26*, 351–357.
Wang, Q. J. (1991), The POT model described by the generalized Pareto distribution with Poisson arrival rate, *J. Hydrol.*, *129*, 263–280.
Wasko, C., and A. Sharma (2015), Steeper temporal distribution of rain intensity at higher temperatures within Australian storms, *Nat. Geosci.*, *8*, 527–529, doi:10.1038/ngeo2456.
Wasko, C., A. Sharma, and F. Johnson (2015), Does storm duration modulate the extreme precipitation-temperature scaling relationship?, *Geophys. Res. Lett.*, *42*, 8783–8790, doi:10.1002/2015GL066274.
Werner, P. C., and F. W. Gerstengarbe (1997), Proposal for the development of climate scenarios, *Clim. Res.*, *8*, 171–182.
Willems, P. (2000), Compound intensity/duration/frequency relationships of extreme precipitation for two seasons and two storm types, *J. Hydrol.*, *233*, 189–205.
Zhang, J. (2007), Likelihood moment estimation for the generalized Pareto distribution, *Aust. N. Z. J. Stat.*, *49*(1), 69–77.