Preprint manuscript No.
(will be inserted by the editor)

Comparing models for the exceedances over high thresholds with applications to extreme temperatures

Andrés M. Alonso · **Patricia de Zea Bermudez** ·
Manuel G. Scotto.

Abstract In this paper, a subsampling-based procedure for the comparison of the exceedances distributions of stationary time series is proposed. To this extent, a test based on distances between Generalized Pareto distributions is introduced and studied in some detail. The performance of the testing procedure is illustrated through a simulation study and with an empirical application to a set of data concerning daily maximum temperature in seventeen regions of Spain.

Keywords Extreme value theory · generalized Pareto distribution · peaks over threshold · subsampling

1 Introduction

The Generalized Pareto Distribution (GPD, in short) plays a central role in extreme value statistics as a distribution of the sample of excesses above a high threshold,

The second author would like to thank the support of the Department of Statistics of the University Carlos III, Madrid (Spain), where she stayed during her last four months of Sabbatical Leave. This stay was financially supported by the grant SFRH/BSAB/1138/2011 and by the project PEst-OE/MAT/UI0006/2011, from Fundação Nacional para a Ciência e Tecnologia - FCT, Portugal. The first author also acknowledges the support of CICYT (Spain) grants SEJ2007-64500 and ECO2011-25706. The third author acknowledges the support of grant PTDC/MAT/ 118335/2010.

Andrés M. Alonso
Department of Statistics and Instituto Flores de Lemus
Universidad Carlos III de Madrid, Spain
E-mail: amalonso@est-econ.uc3m.es

Patricia de Zea Bermudez
Department of Statistics and Operations Research
Universidade de Lisboa, Portugal
E-mail: pcbermudez@fc.ul.pt

Manuel G. Scotto
Department of Mathematics
Universidade de Aveiro, Portugal
E-mail: mscotto@ua.pt

Accepted at Stochastic Environmental Research and Risk Assessment

https://doi.org/10.1007/s00477-013-0809-8

method commonly referred to as the Peak-Over-Threshold (POT, in short) method (Davison and Smith, [9]; Pickands, [35]). The POT method has been extensively used in hydrology (e.g. Katz et al., [22]), finance and insurance (e.g., Gupta and Liang, [20]; Chavez-Demoulin and Embrechts, [7]; Lauridsen, [24]), precipitation (e.g., Furrer and Katz, [18]; Cameron et al., [3]), environment (e.g., Draghicescu and Ignaccolo, [14]; Mendes et al. [28]; Nogaj et al, [33]; Tobías and Scotto, [40]; Naveau et al., [31]; Niu, [32]), and ocean engineering (e.g., Letetrel et al. [26]; Menendez et al. [30]; Méndez et al. [29]) just to mention a few; see Scarrott and MacDonald [37] for further applications. In the last two decades extensive research has been focused on characterizing the GPD and on deriving probabilistic and statistical results; see, for example, de Zea Bermudez and Kotz ([12],[13]), de Haan and Ferreira [11], Castillo et. al. [6], and Kotz and Nadarajah [23] for details.

A typical research question of many empirical studies in the area of environmental processes attempts comparing differences among time series with regard to their corresponding extremal behavior. The study of extreme values is of particular interest in a climate change context. For example, in studies of regional variability of daily mean temperature time series it is important to identify locations exhibiting similar behavior in terms of their corresponding distributions of return values for a pre-defined period of time (see e.g., Scotto et al. [39]). Further examples can be found in the analysis of regional variability of tide gauge records. Knowledge about extreme sea-levels is essential for prediction of flooding risks, coastal management and design of coastal infrastructure systems; see Scotto et al. [38] and the references therein.

A natural way of inferring on such differences should be to employ statistical tests of equality of Generalized Pareto distributions (GPDs). More precisely, let $(X_t) \equiv (X_t : t \in \mathbb{Z})$ and $(Y_t) \equiv (Y_t : t \in \mathbb{Z})$ be two strictly stationary processes with underlying models F_X and F_Y, respectively. In addition, assume that G_X and G_Y are the generalized Pareto distributions for excesses above a sufficiently high threshold of (X_t) and (Y_t), respectively. The interest lays on testing the null hypothesis that

$$H_0 : G_X(x) = G_Y(x) \text{ for almost all } x \in \mathbb{R}^+, \tag{1}$$

against the alternative hypothesis H_1 that $G_X(x) \neq G_Y(x)$ on a set with positive measure.

To this extent, in this paper a testing procedure based on subsampling techniques for testing the equality of GPDs related with the excesses of two stationary time series, which may not be considered independently generated, is introduced. The Kolmogorov-Smirnov distance and the L_1-Wasserstein distance are adopted as a metric between the two corresponding empirical distribution functions. We follow closely the procedure introduced by Alonso and Maharaj [1]. Their ideas will be applied throughout the paper. The performance of the testing procedure is illustrated through a simulation study and with an empirical application to a set of data concerning daily mean temperature collected over the period 1990 and 2004 in 17 locations in Spain. Further results concerning extreme temperatures in Spain can be found in Furió and Meneu [17], Fernández-Montes and Rodrigo [10], Brunet et al. [4] and

García-Herrera et al. [19]

The rest of the paper is organized into the following sections: Section 2 describes the methodological aspects concerning extreme value theory (EVT, in short) and hypothesis testing based on subsampling. Section 3 presents the results of an extensive Monte Carlo study and Section 4 illustrates the behavior of the proposed procedure with a real data example. A brief summary of conclusions is given in Section 5.

2 Methodological aspects

2.1 Basic concepts about EVT

The Generalized Extreme Value (GEV, in short) distribution as well as the GPD, are two of the traditional distributions used for modeling extremal events. The GEV distribution is obtained as the limiting distribution of the maxima, conveniently normalized, of a set of independent and identically distributed (i.i.d.) random variables (r.v's) $X_1, X_2, \ldots, X_n$ with common cumulative distribution function (c.d.f.) F. Effectively, classical extreme value theory states that if sequences of real numbers a_n ($a_n > 0$) and b_n exist for all $n \in \mathbb{N}$ such that

$$\lim_{n\to\infty} P\left(\frac{\max(X_1, X_2, \ldots, X_n) - b_n}{a_n} \leq x\right) = \lim_{n\to\infty} F^n(a_n x + b_n) = G(x), \qquad (2)$$

for all the values x for which G is a continuous function, then F is said to belong to the max-domain of attraction of G, commonly denoted by $F \in D(G)$. The c.d.f. $G(.)$ is the GEV(k, μ, σ) and is given by

$$G(x \mid k, \mu, \sigma) = \begin{cases} \exp\left[-\left(1 + k\frac{x-\mu}{\sigma}\right)^{-\frac{1}{k}}\right], \; 1 + k\frac{x-\mu}{\sigma} > 0, \quad k \neq 0 \\ \exp\left[-\exp\left(-\frac{x-\mu}{\sigma}\right)\right], \quad x \in \mathbb{R}, \quad k = 0 \end{cases}, \qquad (3)$$

where $k \in \mathbb{R}$, $\mu \in \mathbb{R}$ and $\sigma > 0$ are the shape, location and scale parameters, respectively. The Fréchet, Gumbel and Weibull domains of attraction are obtained for $k > 0$, $k = 0$ and $k < 0$, respectively. The Fréchet domain of attraction embraces heavy-tailed distributions with polynomially decaying tails. All d.f.'s belonging to Weibull domain of attraction are light-tailed and have finite right endpoints. The intermediate case $k = 0$ is of particular interest in many applied sciences, not only because of the simplicity of inference within the Gumbel domain of attraction, but also for the great variety of distributions possessing an exponential tail, whether having finite right endpoint or not.

An alternative and more efficient approach to modeling the sample of maxima is the POT methodology. The POT was developed by the hydrologists and consists of fitting a parametric model to the excesses (or to the exceedances) above a sufficiently high threshold. In this framework, let G_u be the conditional c.d.f. of the excesses above u, defined as

$$G_u(x) = P(X-u \leq x \mid X > u) = \frac{G(u+x)-G(u)}{1-G(u)}, \quad 0 \leq x \leq x_G - u.$$

Pickands [35], and also Balkema and de Haan [2], proved that:

$$\lim_{u \to u_G} \sup_{0<x<x_G-u} |G_u(x) - G(x \mid k,\sigma)| = 0, \tag{4}$$

being x_G ($x_G \leq \infty$) the endpoint of the distribution and G the c.d.f. of the two-parameter GPD given by:

$$G(x \mid k,\sigma) = \begin{cases} 1-\left(1+\frac{kx}{\sigma}\right)^{-1/k}, & k \neq 0 \\ 1-\exp(-\frac{x}{\sigma}), & k=0 \end{cases}, \tag{5}$$

where k and σ are the shape and scale parameters, respectively. For $k \geq 0$, $x > 0$, while $0 < x < -\sigma/k$ provided that $k < 0$. Like the GEV distribution the GPD has been widely applied to several areas such as environment, finance, forest fires and climatology. It is relevant to point out that both GEV distribution and the GPD, although resulting from different approaches share the same shape parameter.

In real data applications, the choice of an adequate threshold is frequently a very difficult task. The use of the Mean Excess Function (MEF, in short) is sometimes useful for determining the value u such that

$$P(X > x+u \mid X > u) \approx \left(1+\frac{kx}{\sigma}\right)^{-1/k}.$$

The theoretical MEF is defined as $e(u) := E(X-u \mid X > u)$ and is estimated by

$$e_n(u) := \frac{\sum_{i=1}^n (x_i-u)I(x_i > u)}{\sum_{i=1}^n I(x_i > u)},$$

where $x_1, x_2, \ldots, x_n$ is an observed sample and $I(\cdot)$ the indicator function. For a GPD(k,σ), the MEF is given by

$$e(u) := \frac{\sigma}{1-k} + u\frac{k}{1-k}, \; k < 1 \text{ and } \sigma + ku > 0. \tag{6}$$

If the data fits to a GPD then the plot of the empirical MEF, as a function of u, should be (approximately) a straight line above the level u. The slope and intercept of the straight line take the form $k/(1-k)$ and $\sigma/(1-k)$, respectively. For $k=0$, the empirical MEF is constantly equal to σ, whereas it increases (or decreases) for $k > 0$ (or $k < 0$). One of the benefits of utilizing the GPD, besides the more profitable use of the information contained in the sample (as compared to using the GEV for modeling the sample maxima), is the so-called stability property. Indeed, if $X \sim$GPD(k,σ) then $X-u \mid X > u \sim$GPD$(k, \sigma + ku)$. This property is very useful in the choice of an adequate threshold u. Another common approach for locating an appropriate u

consists on plotting the estimates of the shape parameter k as a function of u and try to locate a stable part in the graph, where there seems to be a balance between bias and variance. A simpler and most useful alternative consists on considering the threshold as the 0.90 sample quantile of the distribution, as indicated by DuMouchel [15].

Several methods have been proposed in the literature for estimating the parameters of the GPD. De Zea Bermudez and Kotz ([12],[13]) reviewed several of these methods (see also Mackay et al. [27] for un update). The maximum likelihood stands out as the most widely used estimation method for the GPD, along with the probability weighted moments and the method of moments. Another very useful method is the elemental percentile method proposed by Castillo and Hadi [5]. This method is easy to apply, never produces non-feasible estimates and it is an alternative to the three methods referred to above in many circumstances, namely when the algorithm used for producing the maximum likelihood estimates fails to converge or when the sample size is very small.

It is worth mentioning that the methods explained in the previous paragraphs rely on the fact that the sample maxima or the sample of excesses above u are i.i.d. This is valid in a limited number of situations but definitely not when considering an extreme value analysis within a time series framework. Essentially, as long as the time series is stationary and the dependence between the large observations decreases, as the distance between the observations increases, then the general i.i.d. theory applies after some adaptations. Basically, it is sufficient to admit that the stationary time series satisfies Leadbetter's-$D(u_n)$ condition (see Leadbetter et al. [25] for details).

The dependency of the extremes is expressed by means of the extremal index, usually denoted by θ. The extremal index satisfies $0 \leq \theta \leq 1$. For i.i.d. r.v's $\theta = 1$. Nonetheless, this is only a necessary condition for independence. As the value of θ decreases, the dependency of the largest observations increases and as such, the exceedances tend to form clusters. Effectively, the limiting mean size of the clusters is given by $1/\theta$. In spite of the clustering behavior of the largest observations, the limiting distributions of the samples of maximums of the stationary and of the i.i.d. sequences belong to the same GEV family (see, e.g., Coles [8] for further details).

The dependency of the largest observations is also an issue to address when using the POT approach. This problem is tackled by making use of a so-called declustering-method. One of the most common techniques (see, e.g. Coles, [8]) consists of the following steps:

- define what a cluster exceedance consists on;
- based on the observed sample, determine the existing clusters;
- select the maximum value of each cluster;
- fit a GPD to the sample of cluster maxima (considering that the maximum cluster values are independent).

This is the declustering-method applied in the data analysis presented in Section 4.

2.2 Hypothesis testing procedure based on subsampling

Let (X_t) and (Y_t) be two strictly stationary processes with underlying models P_X and P_Y, respectively. Let $\boldsymbol{X} := (X_1, X_2, \ldots, X_{n_X})$ and $\boldsymbol{Y} := (Y_1, Y_2, \ldots, Y_{n_Y})$ be vectors of observations from (X_t) and (Y_t), respectively. Let G_X and G_Y be the generalized Pareto distributions for the excesses above a sufficiently high threshold of (X_t) and (Y_t), respectively.

We focus on the following two-sided problem of testing

$$\begin{array}{l} H_0: \; G_X = G_Y \\ H_1: G_X \neq G_Y \end{array}, \tag{7}$$

i.e., to test if the limiting distribution of the excesses is the same in both processes. To this extent, we consider a distance-based test statistic

$$T_{n,d} := \tau_n d(\widehat{G}_{n,X}, \widehat{G}_{n,Y}), \tag{8}$$

being the normalizing constant $\tau_n := \sqrt{(n_X * n_Y)/(n_X + n_Y)}$, $\widehat{G}_{n_X,X} = GPD(\widehat{k}_X, \widehat{\sigma}_X)$ and $\widehat{G}_{n_Y,Y} = GPD(\widehat{k}_Y, \widehat{\sigma}_Y)$ are the estimated GPD's from $\boldsymbol{X}$ and $\boldsymbol{Y}$ (i.e., $(\widehat{k}_X, \widehat{\sigma}_X)$ and $(\widehat{k}_Y, \widehat{\sigma}_Y)$ are estimated using $\boldsymbol{X}$ and $\boldsymbol{Y}$, respectively), and $d(\cdot,\cdot)$ is an adequate metric capturing the discrepancies between the two estimated distribution functions. In this case, the Kolmogorov-Smirnov distance

$$d_{KS}(F_1, F_2) := \sup_{-\infty < x < +\infty} |F_1(x) - F_2(x)| \tag{9}$$

and the L_1-Wasserstein distance

$$d_W(F_1, F_2) := \int_{-\infty}^{+\infty} |F_1(x) - F_2(x)| \, dx, \tag{10}$$

are adopted.

The general subsampling approach (see, e.g. Politis et al., [36]) is applied to the statistic $T_{n,d}$ as follows:

1. Let $\boldsymbol{X}_j := (X_j, \ldots, X_{j+l-1})$ and $\mathbf{Y}_j := (Y_j, \ldots, Y_{j+l-1})$ with $j = 1, \ldots, n-l+1$ be the subsamples of l consecutive observations from $\boldsymbol{X}$ and $\boldsymbol{Y}$, respectively. The j-th subsampling statistic, $T_{l,d}^{(j)}$ is calculated through

$$T_{l,d}^{(j)} := \tau_l d(\widehat{G}_{l,X_j}, \widehat{G}_{l,Y_j}), \tag{11}$$

   where $\widehat{G}_{l,X_j} = GPD(\widehat{k}_{X_j}, \widehat{\sigma}_{X_j})$ and $\widehat{G}_{l,Y_j} = GPD(\widehat{k}_{Y_j}, \widehat{\sigma}_{Y_j})$ are the estimated GPD's from the subsamples $\boldsymbol{X}_j$ and $\boldsymbol{Y}_j$, respectively.
2. The critical value for the test is obtained as the $1-\alpha$ quantile of $\widehat{\mathscr{H}_{n,l}}(x) := \frac{1}{n-l+1} \sum_{j=1}^{n-l+1} I(T_l^{(j)} \leq x)$ defined as

$$h_{n,l}(1-\alpha) := \inf \left\{ x : \widehat{\mathscr{H}_{n,l}}(x) \geq 1-\alpha \right\}. \tag{12}$$

3. Thus, H_0 is rejected if and only if $T_{n,d} > h_{n,l}(1-\alpha)$.

Notice that the proposed algorithm remains valid for dependent series since the subsamples $(\boldsymbol{X}_j, \boldsymbol{Y}_j)$ can be considered as a vector of size l from the bidimensional process $\{(X_t, Y_t)\}$.

3 Simulation study

In this section, the behavior of the proposed testing procedure for finite samples is analyzed using data sets generated from the autoregressive model

$$X_t = \phi X_{t-1} + Z_t,\ t \in \mathbb{Z}, \tag{13}$$

where (Z_t) are i.i.d. innovations. The simulation study contemplates two different values for ϕ, namely 0 and 0.5, and three possible distribution for Z_t, namely (a) Uniform$(0,\theta)$, (b) Exponential(θ) and (c) Fréchet(θ) are considered with the set of parameters $\theta = \{1, 1.25, 1.5, 2, 5\}$. These three distributions exhibit different patterns in what concerns to their extremal behavior. In the case of uniform and exponentially distributed innovations the test based on L_1-Wasserstein distance is considered. In the case of Fréchet distributed innovations the test based on Kolmogorov-Smirnov distance is adopted since the Fréchet distribution can exhibit infinite first moment depending on the value of its corresponding tail index. Notice that L_1-Wasserstein distance metrizes weak convergence plus convergence of first absolute moments. A sample size $n = 2048$ and the nominal size $\alpha = 0.05$ is adopted to carry out the simulation analysis. We present the simulations results for pairs of series that are assumed to be independently generated from each other but using the algorithm based on the expression (12).

The threshold u is fixed as the 95% sample quantile. The fact that sometimes the sample size of excesses above u is small justifies the use of the Elemental Percentile Method (EPM), proposed by Castillo and Hadi [5] for estimating the parameters of the corresponding GPDs, as an alternative to the celebrated ML method. The constraints imposed by the ML method in terms of the regularity conditions, as well as the issues associated to the convergence of the algorithms used to produce the parameter estimates, may cause problems to a simulation study. The EPM is known to be a very flexible and easy to use method that never produces unfeasible results (see e.g. de Zea Bermudez and Kotz [12]).

In Tables 1 and 2 below the estimated size of the test for $l = 256$, 512 and 1024 are displayed. These values of l were selected to assess the sensitivity of the test to a reasonable range of subsampled lengths. To estimate the size of the test 1000 independent replicas are generated. Therefore, an estimated size between 0.0365 and 0.0635 would not be significantly different from the nominal value 0.05. As expected, the results strongly dependent on the subsampled lengths. While reasonable size estimates are obtained, for the lower values of l (i.e., $l = 256$ and 512), in several cases they are significantly different from the nominal size of 0.05. A possible solution to obtain sizes closer to the nominal size is to use a calibration procedure as in Section 9.4 of Politis et al. [36].

Table 1 Estimated size of the test ($n = 2048$) applied to the AR model in (13) with uniformly and exponentially distributed innovations.

	Uniform distribution				Exponential distribution			
$\phi = 0.0$	θ	$l = 256$	$l = 512$	$l = 1024$	θ	$l = 256$	$l = 512$	$l = 1024$
	1.00	0.062	0.098	0.088	1.00	0.050	0.089	0.089
	1.25	0.071	0.080	0.087	1.25	0.050	0.080	0.083
	1.50	0.065	0.093	0.097	1.50	0.054	0.080	0.100
	2.00	0.088	0.100	0.098	2.00	0.063	0.092	0.087
	5.00	0.076	0.095	0.095	5.00	0.063	0.104	0.106
$\phi = 0.5$	θ	$l = 256$	$l = 512$	$l = 1024$	θ	$l = 256$	$l = 512$	$l = 1024$
	1.00	0.063	0.094	0.097	1.00	0.028	0.075	0.075
	1.25	0.069	0.082	0.085	1.25	0.031	0.075	0.078
	1.50	0.061	0.092	0.095	1.50	0.030	0.063	0.079
	2.00	0.070	0.085	0.091	2.00	0.030	0.079	0.086
	5.00	0.065	0.095	0.083	5.00	0.036	0.094	0.079

Table 2 Estimated size of the test ($n = 2048$) applied to the AR model in (13) with Fréchet distributed innovations.

$\phi = 0.0$	θ	$l = 256$	$l = 512$	$l = 1024$
	1.00	0.027	0.099	0.112
	1.25	0.048	0.130	0.123
	1.50	0.054	0.137	0.122
	2.00	0.069	0.135	0.112
	5.00	0.097	0.149	0.132
$\phi = 0.5$	θ	$l = 256$	$l = 512$	$l = 1024$
	1.00	0.036	0.047	0.095
	1.25	0.026	0.070	0.106
	1.50	0.017	0.080	0.064
	2.00	0.029	0.086	0.052
	5.00	0.057	0.115	0.056

Tables 3 and 4 show the estimates for the power of the following hypothesis test:

$$\begin{array}{l} H_0 : \ G_X = G_Y \\ H_1 : G_X \neq G_Y \end{array}, \tag{14}$$

where G_X corresponds to time series generated using the AR model in (13) with error distribution Uniform$(0, \theta = 1)$ (or Exponential$(\theta = 1)$ or Fréchet$(\theta = 1)$) and G_Y corresponds to time series generated by the same model with error distribution Uniform$(0, \theta)$ (or Exponential(θ) or Fréchet(θ)) with $\theta \geq 1$. As expected, the power increases as θ increases since the models considered in the null and the alternative hypothesis behave very different. For largest values of θ, the power approaches 1.

Table 3 Estimated power of the test ($n = 2048$) applied to the AR model in (13) with uniformly and exponentially distributed innovations.

	Uniform distribution				Exponential distribution			
$\phi = 0.0$	θ	$l = 256$	$l = 512$	$l = 1024$	θ	$l = 256$	$l = 512$	$l = 1024$
	1.00	0.062	0.098	0.090	1.00	0.053	0.090	0.088
	1.25	1.000	0.998	0.977	1.25	0.209	0.268	0.223
	1.50	1.000	1.000	0.999	1.50	0.490	0.518	0.413
	2.00	1.000	1.000	1.000	2.00	0.836	0.731	0.544
	5.00	1.000	1.000	1.000	5.00	0.867	0.851	0.741
$\phi = 0.5$	θ	$l = 256$	$l = 512$	$l = 1024$				
	1.00	0.063	0.094	0.097	1.00	0.026	0.076	0.075
	1.25	0.580	0.523	0.403	1.25	0.135	0.179	0.160
	1.50	0.773	0.684	0.545	1.50	0.291	0.294	0.238
	2.00	0.814	0.717	0.578	2.00	0.646	0.530	0.357
	5.00	0.865	0.774	0.607	5.00	0.716	0.664	0.454

Table 4 Estimated power of the test ($n = 2048$) applied to the AR model (13) with Fréchet distributed innovations.

$\phi = 0.0$	θ	$l = 256$	$l = 512$	$l = 1024$
	1.00	0.028	0.100	0.113
	1.25	0.584	0.434	0.401
	1.50	0.995	0.924	0.740
	2.00	1.000	1.000	0.981
	5.00	1.000	1.000	1.000
$\phi = 0.5$	θ	$l = 256$	$l = 512$	$l = 1024$
	1.00	0.036	0.046	0.095
	1.25	0.881	0.787	0.660
	1.50	0.999	0.988	0.947
	2.00	1.000	1.000	0.998
	5.00	1.000	1.000	0.991

4 Application to a real data set

In this section, an empirical application to time series is presented. The data consist of the daily maximum temperatures observed in each of 17 autonomous communities of Spain, from 1990 to 2004 (*Ceuta* and *Melilla* are autonomous cities and were not considered in this study). The temperatures exhibit the usual yearly seasonal variation, although there is no apparent trend across the years. As an example, the temperatures observed in the *Comunidad de Madrid* are plotted in Figure 1. The rest of the regions exhibit similar patterns. Having in mind that the aim of this study is to compare the largest values of the time series, one of the options is to model the summer periods and disregard the temperatures recorded in the other seasons of the year. However, this procedure was not followed here because it jeopardizes the dependence structure within the series, in the sense that we end up by modeling a set of disjoint summer periods. Notice that if we consider disjoint summer periods as the complete series then that series will have a spurious correlation of order equal to a quarter of a year. Moreover, large temperatures might also be observed in late spring and in the first days of the autumn seasons. An alternatively and more reasonable approach is to consider all the temperatures that exceed a very high threshold u. As early mentioned, the choice of the 90th percentile is very common (see DuMouchel, 1983) and is usually accurate enough. However, the fact that the exceedances are determined irrespectively of the time period requires that the threshold is slightly raised, otherwise temperature not so “extreme” might be included. As such, the 95th percentile was chosen instead. Notice that taking a hight percentile as threshold is almost equivalent to the approach based on modeling summer periods as proposed by Coles (2001). The $P_{0.95}$ is also plotted in Figure 1. A careful analysis of this sample reveals that approximately 97% of the observations were recorded during summer days and the 3% (7 out of 263) remaining were recorded in last days of spring and none was observed in autumn days. The sample corresponds to approximately 19% of the summer days of the time frame considered in this study.

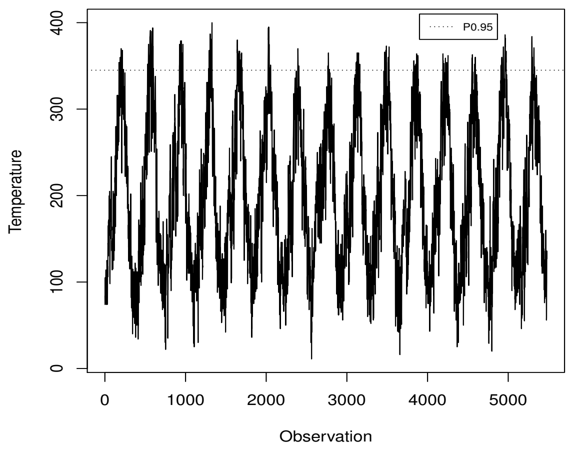

Fig. 1 Daily maximum temperatures observed in the *Comunidad de Madrid* and the corresponding $P_{0.95}$

The plots of the sample MEF can also be used for justifying the choice of the sample 95th percentile. As mentioned before, the threshold should be the temperature onwards it is reasonable to model the data by a GPD. The plots for the *Cantabria*, *Comunidad de Madrid* and *Región Murciana* are presented in Figure 2. The thresholds considered are quite reasonable for the first and third regions. In fact, the linearity starts to be evident from the $P_{0.95}$ onwards. The plot for the *Comunidad de Madrid* is very difficult to analyze. Therefore, virtually any high threshold could be chosen for this data set. The box-plots of the exceedances above the data of each region, presented in Figure 3, clearly show that *Cantabria* and *Asturias* stand out as the two regions with milder temperatures and *Andalucía*, *Castilla La-Mancha* and *Extremadura* as the most extreme, as would be expected. Table 5 shows the dates of the consecutive days with temperatures above the threshold for the *Comunidad de Madrid* data for 2000-2004. It is clear that half of the groups have size 2 and 79% have sizes less or equal to 4. The temperatures observed in 2003 were specially high

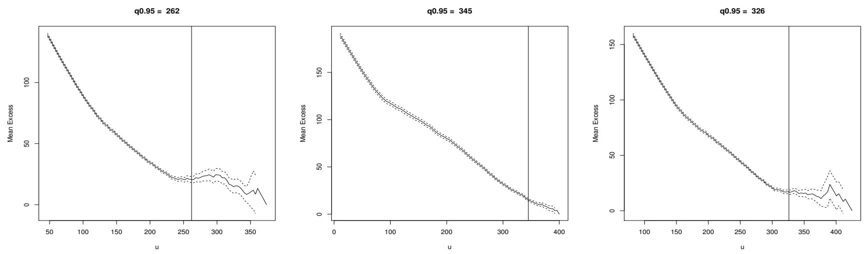

Fig. 2 Sample MEF for the daily maximum temperatures for *Cantabria* (left), the *Comunidad de Madrid* (centre) and the *Regiǿn Murciana* (right).

Fig. 3 Box-plot of the exceedances above the 0.95 percentile for each community.

in this community. Table 5 shows that the exceedances above the $P_{0.95}$ are not in-

Table 5 Sizes of the groups of consecutive hot temperatures for the *Comunidad de Madrid* sample for 2000-2004 (* hot day = temperature above $P_{0.95}$).

Beginning of the period	Size of group	Beginning of the period	Size of group
15-08-2000	3	21-06-2003	2
21-06-2001	4	08-07-2003	5
01-07-2001	2	30-07-2003	9
23-08-2001	2	09-08-2003	7
16-06-2002	2	26-06-2004	6
22-06-2002	2	04-07-2004	2
26-07-2002	4	22-07-2004	4
		31-07-2004	2
Year	**Number of hot days***	**Year**	**Number of hot days***
2000	9	2003	25
2001	11	2004	15
2002	9		

dependent as required by the classical EVT. As previously mentioned the extremal index, θ, measures the degree of clustering of the large observations. It is generally estimated by n_c/n_u, where n_c is the number of clusters and n_u represents the number of exceedances above u. The estimates of θ for the 17 communities of Spain obtained for $r = 2,3$ and 4 are presented in Table 6. This table also contains the estimates of the parameters of the GPD, k and σ, estimated by the EPM (see Castillo and Hadi [5]). The numbers in brackets are the corresponding standard errors computed using 1000 bootstrap samples. The figures presented show that neither the estimates of k and σ nor the estimates of θ seem to be significantly affected by the different values of r. Therefore, it was decided to consider $r = 2$. In this situation, two observations above the u are considered to belong to different clusters if there are, at least, two consecutive observations below u. The estimates of θ represent roughly 2 to 3 very hot consecutive days. The estimates of k are all negative reflecting underlying light-tailed distributions with a finite upper-bound $x_G < \infty$. When attention focuses on estimating

Table 6 Estimates of θ, k and σ for the 17 Spanish communities (standard errors of the estimates between brackets)

Community	u	$r=2$			$r=3$			$r=4$		
		$\hat{\theta}$	$\hat{k}$	$\hat{\sigma}$	$\hat{\theta}$	$\hat{k}$	$\hat{\sigma}$	$\hat{\theta}$	$\hat{k}$	$\hat{\sigma}$
Andalucía	394.0	0.295	-0.272	28.1	0.276	-0.256	27.6	0.239	-0.268	28.7
			(0.1336)	(4.2560)		(0.1351)	(4.2771)		(0.1441)	(4.7911)
Aragón	359.1	0.336	-0.244	24.0	0.285	-0.257	25.1	0.234	-0.388	31.9
			(0.1712)	(3.5106)		(0.1877)	(4.2721)		(0.2071)	(4.9889)
Asturias	258.1	0.558	-0.186	30.0	0.496	-0.203	31.4	0.442	-0.234	33.9
			(0.0538)	(3.2348)		(0.0577)	(3.6246)		(0.0639)	(4.0917)
Cantabria	262.0	0.538	-0.103	28.8	0.492	-0.116	30.1	0.455	-0.144	32.3
			(0.0730)	(3.5278)		(0.0778)	(3.9763)		(0.084)	(4.4241)
Cast. La Mancha	375.1	0.303	-0.359	26.6	0.266	-0.406	29.0	0.248	-0.421	29.8
			(0.1453)	(3.4944)		(0.1601)	(4.1325)		(0.1665)	(4.3173)
Cast. y León	344.0	0.312	-0.367	30.2	0.304	-0.346	29.2	0.274	-0.402	32.3
			(0.1464)	(4.3082)		(0.1505)	(4.3377)		(0.1571)	(4.5593)
C. de Madrid	345.0	0.297	-0.436	28.2	0.259	-0.486	30.7	0.232	-0.474	30.4
			(0.1046)	(4.0121)		(0.1065)	(4.0892)		(0.1217)	(4.8604)
C. Valenciana	326.1	0.518	-0.101	25.3	0.442	-0.146	28.6	0.391	-0.146	29.2
			(0.0912)	(3.0190)		(0.0908)	(3.4058)		(0.1014)	(3.8674)
Cataluña	355.0	0.328	-0.356	25.5	0.302	-0.409	28.0	0.240	-0.432	29.5
			(0.1349)	(3.4756)		(0.1438)	(3.7703)		(0.1597)	(4.1525)
Extremadura	376.0	0.335	-0.390	34.0	0.323	-0.408	35.1	0.292	-0.407	35.3
			(0.1130)	(4.1356)		(0.116)	(4.3030)		(0.1188)	(4.2967)
Galicia	344.0	0.377	-0.281	31.7	0.343	-0.321	34.3	0.317	-0.305	33.7
			(0.0777)	(4.2125)		(0.0848)	(4.9030)		(0.0899)	(4.9676)
I. Baleares	314.0	0.453	-0.076	16.5	0.362	-0.128	19.2	0.299	-0.162	21.1
			(0.0645)	(2.1368)		(0.0788)	(2.8508)		(0.0853)	(3.2580)
I. Canarias	306.0	0.474	-0.027	20.0	0.425	-0.041	21.1	0.395	-0.070	22.9
			(0.0741)	(2.4666)		(0.0744)	(2.9475)		(0.0822)	(3.1441)
La Rioja	342.0	0.354	-0.519	36.7	0.300	-0.536	38.0	0.277	-0.535	38.1
			(0.0889)	(4.7235)		(0.1033)	(5.4904)		(0.1079)	(5.7174)
Navarra	330.0	0.409	-0.484	40.0	0.367	-0.502	41.3	0.328	-0.538	43.8
			(0.0949)	(4.4548)		(0.1104)	(4.9934)		(0.1159)	(5.5313)
País Vasco	314.0	0.522	-0.269	38.3	0.448	-0.314	42.4	0.411	-0.332	44.0
			(0.1305)	(4.3443)		(0.1418)	(4.9677)		(0.1494)	(5.3471)
Reg. Murciana	358.0	0.510	-0.079	24.4	0.432	-0.109	26.8	0.382	-0.102	26.9
			(0.1546)	(2.9362)		(0.1686)	(3.7120)		(0.1777)	(4.0164)

extreme quantiles then the aim is to search for a value x such that $P(X > x) = p$, for a fixed and very small probability p. If $p = 1/m$, then x is exceeded one in every m observations (m should be a large number). In some occasions it is more interesting to express the extreme quantiles in terms of N-year return levels. Generally $m = n_y N$, where $n_y = 365$ is the number of observations per year and N is the number of years. The extreme quantiles are computed as

$$x_m = u + \frac{\sigma}{k}\left[(m\tau_u\theta)^k - 1\right],$$

where $\tau_u\theta$ is estimated by n_c/n, being n_c the number of clusters above the threshold u. Some very extreme quantiles are presented in Table 7. The probabilities presented in Table 7 represent, approximately, the occurrence of a very hot day once every 50 and 100 years, respectively. The results contained in Table 7 show that, for some communities, the temperatures that are expected to be exceeded in a time window of 50-years are very similar to the ones observed for 100-years. *Comunidad de Madrid*, *La Rioja* and *Navarra* are examples of such situations. In the overall, for the 17 communities, the average increase in temperature in 50 years is about half degree Celsius. However, comparing the maximum temperatures (second column in Table 7) observed and the $50-$years return level, the average increase is about 1 degree Celsius, which is a considerable augmentation.

Table 7 $N = 50$ and $N = 100$ return levels for the 17 Spanish communities.

Autonomous Community	Maximum temperature observed	$p = 0.000055$ $m = 18250$ $N = 50$	$p = 0.000027$ $m = 36500$ $N = 100$
Andalucía (AND)	466	475	479
Aragón (ARA)	425	433	437
Asturias (AST)	356	369	375
Cantabria (CAN)	374	393	404
Castilla-La Mancha (CLM)	434	439	441
Castilla y León (CyL)	410	416	418
Comunidad de Madrid (MAD)	400	404	405
Comunidad Valenciana (VAL)	425	442	451
Cataluña (CAT)	412	417	419
Extremadura (EXT)	448	454	456
Galicia (GAL)	426	435	439
Islas Baleares (IBA)	380	393	400
Islas Canarias (ICA)	397	418	429
La Rioja (RIO)	406	409	410
Navarra (NAV)	404	408	409
País Vasco (EUS)	419	429	434
Región Murciana (MUR)	457	476	486

Finally, the proposed testing procedure is applied to test whether their exceedances distributions are equal or not. In Figure 4, we summarize the results in a dendrogram based on complete linkage and (1-p)-values as a dissimilarity measure. Notice that the joins above 0.95 (or above $1-0.05/136$, with 136 representing the total number of comparisons, if the Bonferroni multiple testing correction is considered) correspond to a test where the null hypothesis is rejected. This cutoff of 0.95 (or $1-0.05/136$) allows to define groups of series having similar distributions. The results using the cutoff 1-0.05/136 are represented in Figure 5. Four clusters are obtained: (C1) Cantabria (CAN), Asturias (AST), País Vasco (EUS) and Navarra (NAV) which have an exceedances mean of 33.0; (C2) Castilla y León (CyL), Comunidad de Madrid (MAD), Galicia (GAL) and La Rioja (RIO) with an exceedances mean of 36.3; (C3) Cataluña (CAT), Extremadura (EXT), Aragón (ARA) and Islas Canarias (ICA) with an exceedances mean of 37.6; (C4) Andalucía (AND), Islas Baleares (IBA), Castilla-La Mancha (CLM), Comunidad Valenciana (VAL) and Región Murciana (MUR) with an exceedances mean of 38.7. From Figure 4 and Figure 5, we derive conclusions that are coherent with the previous descriptive analysis.

5 Conclusions

In this work, a procedure based on subsampling techniques for testing the equality of GPDs related with the excesses of two stationary time series that may not be considered independently generated, has been introduced and applied to a set of daily mean temperature collected at 17 locations in Spain. The clustering results reflect spatial consistency. Furthermore, the analysis identifies a clear distinction between the four northeast communities on the shores of the Bay of Biscay and the remaining communities. A clear distinction is also found between AND, IBA, CLM, VAL and MUR

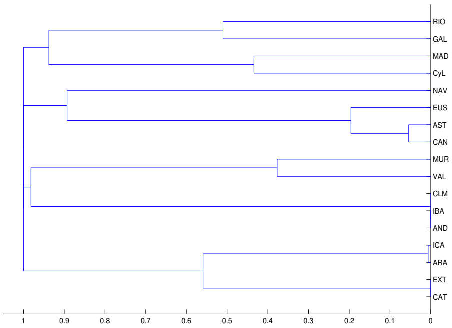

Fig. 4 Dendrogram using complete linkage for (1-p)-values associated with comparison test.

Fig. 5 Cluster defined by the similarity of exceedances distributions.

(being the communities exhibiting highest values) from the remaining locations. Beyond the analysis of temperatures, the inclusion of covariates into the parameters of the GPD in a regression-like approach allowing for trends or cycles in the upper tail, would be also relevant for classification purposes. We might consider the possibility of also having a time and/or space varying threshold for modeling the largest annual temperatures (see, for instance, the recent paper by Northrop and Jonathan [34]). This remains a topic of future research.

References

1. Alonso, A.M., Maharaj, E.A., Comparison of time series using subsampling. Comput. Statist. Data Anal. 50, 2589–2599 (2006)
2. Balkema, A. and de Haan, L., Residual life time at great age. Ann. Prob. 2, 792–804 (1974).
3. Cameron, D., Bevin, K., Tawn, J., Modelling extreme rainfalls using a modified pulse Bartlett-Lewis stochastic rainfall model (with uncertainty). Adv. Water Resour. 24, 203–211 (2001).
4. Brunet, M., Jones, P.D., Sigro, J., Saladié, O., Aguilar, E., Moberg, A., Della-Marta, P.M., Lister, D., Walther, A., Lopez, D., Temporal and spatial temperature variability and change over Spain during 1850-2005. J. Geophys. Res. 112: D12117 (2007).
5. Castillo, E., Hadi, A.S., Fitting the Generalized Pareto Distribution to data. J. Am. Statist. Ass. 92, 1609–1620 (1997)
6. Castillo E., Hadi, A.S., Balakrishnan, N., Sarabia, J.M., Extreme Value and Related Models with Applications in Engineering and Science, pp. 362. Wiley, New Jersey (2004)
7. Chavez-Demoulin, V., Embrechts, P., Smooth extremal models in finance and insurance. J. Risk Insur. 71, 183–199 (2004)
8. Coles, S.G., An Introduction to Statistical Modeling of Extreme Values, pp. 228. Springer-Verlag, London (2001)
9. Davison, A.C., Smith, R.L., Models for exceedances over high tresholds. J. Roy. Statist. Soc, B 52, 393–442 (1990)
10. Fernández-Montes, S., Rodrigo, F.S., Trends in seasonal indices of daily temperature extremes in the Iberian Peninsula, 1929–2005. Int. J. Climatol. (2011) (to appear)
11. de Haan, L., Ferreira, A., Extreme Value Theory - An Introduction, pp. 417. Springer-Verlag, New York (2006)
12. de Zea Bermudez, P., Kotz, S., Parameter estimation of the generalized Pareto distribution - Part I. J. Statist. Plann. Inference 40, 1353–1373 (2010a)
13. de Zea Bermudez, P., Kotz, S., Parameter estimation of the generalized Pareto distribution - Part II. J. Statist. Plann. Inference 40, 1374–1388 (2010b)
14. Draghicescu, D., Ignaccolo, R., Modeling threshold exceedance probabilities of spatially correlated time series. Electron. J. Statist. 3, 149–164 (2009)
15. DuMouchel, W.H., Estimating the stable index α in order to measure tail thickness: a critique. Ann. Statist. 11, 1019–1031 (1983)
16. Embrechts, P., Klüppelberg, C., Mikosch, T., Modelling Extremal Events for Insurance and Finance, pp. 645. Springer-Verlag, Berlin (1997)
17. Furió, D., Meneu, V., Analysis of extreme temperatures for four sites across Peninsular Spain. Theor. Appl. Climatol. 104, 83–99 (2011)
18. Furrer, E.M., Katz, R.W., Improving the simulation of extreme precipitation events by stochastic weather generators. Water Resour. Res. 44, W12439 (2008)
19. García-Herrera, R., Díaz, J., Trigo, R.M., Hernández, E., Extreme summer temperatures in Iberia: health impacts and associated synoptic conditions. Ann. Geophys. 23, 239-–251 (2005)
20. Gupta, A., Liang, B., Do hedge funds have enough capital? A Value at Risk approach. J. Financ. Econ. 77, 219–253 (2005)
21. Grimshaw, S.D., Computing maximum likelihood estimates for the generalized Pareto distribution. Technometrics 35, 185–191 (1993)
22. Katz, R.W., Parlange, M.B., Naveau, P., Statistics of extremes in hydrology. Adv. Water Resour. 25, 1287–1304 (2002)
23. Kotz, S., Nadarajah, S., Extreme Value Distributions: Theory and Applications, pp. 196. Imperial College Press, London (2000)
24. Lauridsen, S., Estimation of Value at Risk by extreme value methods. Extremes 3, 107–144 (2000)
25. Leadbetter, M.R., Lindgren, G. and Rootzén, H. Extremes and Related Properties of Random Sequences and Processes, Springer-Verlag, Berlin (1983)
26. Letretel, C., Marcos, M., Martín-Míguez, B., Woppelmann, G., Sea level extremes in Marseille (NW Mediterranean) during 1885-2008. Cont. Shelf Res. 30, 1267–1274 (2010)
27. Mackay, E., Challenor, P. and AbuBakr, S., A comparison of estimators for the eneralised Pareto distribution, Ocean Engineering, 38, 1338-1346 (2011)
28. Mendes, J.M., de Zea Bermudez, P., Pereira, J., Turkman, K.F., Vasconcelos, M.J.P., Spatial extremes of wild fire sizes: Bayesian hierarchical models for extremes. Environ. Ecol. Statist. 17, 1–28 (2009)

29. Méndez, F.J., Menéndez, M., Luceño, A., Losada, I.J. Analyzing monthly extreme sea levels with a time-dependent GEV model. J. Atmos. Ocean. Technol. 24, 894–911 (2007)
30. Menéndez, M., Méndez, F.J., Izaguirre, C., Losada, I.J., The influence of seasonality on estimating return values of significant wave height. Coast. Eng. 56, 211–219 (2009)
31. Naveau, P., Nogaj, M., Ammann, C., Yiou, P., Cooley, D., Jomelli, V., Statistical methods for the analysis of climate extremes. C. R. Geoscience 337, 1013—1022 (2005)
32. Niu, X.F., Extreme value for a class of non-stationary time series with applications. Ann. Appl. Probab. 7, 508–522 (1997)
33. Nogaj, M., Yiou, P., Parey, S., Malek, F., Naveau, P., Amplitude and frequency of temperature extremes over the North Atlantic region. Geophys. Res. Lett., 33, L10801, doi:10.1029/2005GL024251 (2006)
34. Northrop, P.J., Jonathan, P. Threshold modelling of spatially-dependent non-stationary extremes with application to hurricane- induced wave heights (with discussion), Environmetrics, 22, 799–809 (2011)
35. Pickands, J., Statistical inference using extreme order statistics. Ann. Statist. 3, 119–131 (1975)
36. Politis, D.N., Romano, J.P., Wolf, M., Subsampling. Springer-Verlag, pp. 363. New York (1999)
37. Scarrott, C., MacDonald, A., A review of extreme value threshold estimation and uncertainty quantification. REVSTAT, 10, 33–60 (2012)
38. Scotto, M.G., Alonso, A.M., Barbosa, S.M., Clustering time series of sea levels: extreme value approach. J. Waterway, Port, Coastal, and Ocean Eng. 136, 215–225 (2010)
39. Scotto, M.G., Barbosa, S.M., Alonso, A.M., Extreme value and cluster analysis of European daily temperature series. J. Appl. Statist. 38, 2793–2804 (2011)
40. Tobías, A., Scotto, M.G., Prediction of extreme ozone levels in Barcelona, Spain. Environ. Monit. Assess. 100(1-3), 23–32 (2005)