



# A two-step framework for over-threshold modelling of environmental extremes

P. Bernardara<sup>1,2,5</sup>, F. Mazas<sup>3</sup>, X. Kergadallan<sup>2,4</sup>, and L. Hamm<sup>3</sup>

<sup>1</sup>LNHE, EDF R&D, Chatou, France

<sup>2</sup>Université Paris Est, Saint Venant Laboratory for Hydraulics, ENPC, EDF R&D, CETMEF, Chatou, France

<sup>3</sup>ARTELIA, Grenoble, France

<sup>4</sup>CEREMA, Brest, France

<sup>5</sup>EDF Energy R&D UK Centre, London, UK

*Correspondence to:* P. Bernardara (pietro.bernardara@edf.fr)

Received: 1 August 2012 – Published in Nat. Hazards Earth Syst. Sci. Discuss.: –

Revised: 1 November 2013 – Accepted: 21 November 2013 – Published: 20 March 2014

**Abstract.** The evaluation of the probability of occurrence of extreme natural events is important for the protection of urban areas, industrial facilities and others. Traditionally, the extreme value theory (EVT) offers a valid theoretical framework on this topic. In an over-threshold modelling (OTM) approach, Pickands' theorem, (Pickands, 1975) states that, for a sample composed by independent and identically distributed (i.i.d.) values, the distribution of the data exceeding a given threshold converges through a generalized Pareto distribution (GPD). Following this theoretical result, the analysis of realizations of environmental variables exceeding a threshold spread widely in the literature. However, applying this theorem to an auto-correlated time series logically involves two successive and complementary steps: the first one is required to build a sample of i.i.d. values from the available information, as required by the EVT; the second to set the threshold for the optimal convergence toward the GPD. In the past, the same threshold was often employed both for sampling observations and for meeting the hypothesis of extreme value convergence. This confusion can lead to an erroneous understanding of methodologies and tools available in the literature. This paper aims at clarifying the conceptual framework involved in threshold selection, reviewing the available methods for the application of both steps and illustrating it with a double threshold approach.

## 1 Introduction

A reliable estimation of extreme natural hazard is important for the protection of remarkable natural sites, urban areas, industrial facilities, etc. In particular, extreme natural events include floods, heavy rainfalls, high and low temperatures, strong winds, high sea levels or sea surges, oceanic waves, among many others.

Traditionally, the estimation of the probability of occurrence of such extreme events is performed by fitting a probability distribution to a sample of historical observations for a given phenomenon observed at a given site, usually recorded as a time series of observations. In this framework, the extreme value theory (EVT) (Fréchet, 1928; Gnedenko, 1943; Gumbel, 1958; Pickands, 1975) offers a sound theoretical framework.

In particular, Pickands' theorem (that can be seen as a central limit theorem for extreme values) states that, in a sample composed by independent and identically distributed (i.i.d.) values, the distribution of the data exceeding a given threshold converges towards a generalized Pareto distribution (GPD) (Pickands, 1975). Following this theoretical result, the over-threshold modelling (OTM) approach widely spread in extreme value analyses, together with the application of the GPD, (Davison and Smith, 1990; Simiu and Heckert, 1995; Embrechts et al., 1997; Palutikof et al., 1999; Coles, 2001; Mackay et al., 2001; Pandey et al., 2001; Rosbjerg and Madsen, 2004; Ribatet et al., 2007). It is widely recognized that the choice of the threshold is a critical point

in this approach and the final estimation could significantly depend on its value (Onoz and Bayazit, 2001; Li et al., 2012).

In this theoretical framework, the choice of the appropriate threshold should be a statistical optimization procedure. Given an empirical sample of i.i.d. observations, the selected threshold must be high enough to meet the hypothesis of convergence on the GPD but it should be low enough to limit the variability of the GPD parameter calibration on the sub-sample of observations over the threshold (Beirlant et al., 1996). This is the well-known dilemma between bias and variance.

However, environmental variables are often handled as time series, i.e. discrete realizations of this variable, coming from either observation or modelling, far from being i.i.d. More precisely, environmental time series are often composed by dependent values because of the strong temporal autocorrelation (e.g. Zawadzky, 1987; Smith, 1988; Colombo et al., 1999; Walton, 2000; Marani, 2003; Bernardara et al., 2006). The autocorrelation is indeed explained by the dynamical behaviour of the subjacent physical system and its momentum. Since the pioneering works of Hurst (1951), some studies even suggest that the autocorrelation of some environmental time series could be infinite (Schertzer and Lovejoy, 1997; Elek and Markus, 2004; Koscielny-Bunde et al., 2006).

An attempt to cope at the same time with EVT and autocorrelated data is the introduction of the extremal index, (Leadbetter et al., 1983; Smith and Weissman, 1994; Embrechts et al., 1997; Ancona-Navarrete and Tawn, 2000; Coles, 2001; Beirlant et al., 2004). The extremal index is an extra parameter that allows taking into account data autocorrelation on the extreme value theorem. The extremal index represents the reciprocal of the mean size of event clusters. Estimating this index allow to apply the EVT theorem results directly on a series of auto correlated observations.

However, in general, the EVT cannot be applied directly to the observed data and a data pre-processing is needed in order to build the i.i.d. sample required by its hypothesis. This data pre-processing is often called physical declustering, because it tends to extract independent observations from the time series, which are naturally (physically) clustered.

Moreover, environmental time series can be composed by nonidentically distributed (nonhomogeneous) values. Indeed, natural phenomena can have very different physical genesis, they can exhibit strong seasonality of the observed phenomena or they can depend on other covariates. Among others, Adamowski (2000), Garavaglia et al. (2010) and Allamano et al. (2011) show that mixing heterogeneous samples can lead to biased estimation of extreme value probability of occurrence. Garavaglia et al. (2010) introduced a compound distribution for extreme rainfalls taking into account seasonality and different physical genesis. For wave heights, probability distribution and even time series autocorrelation may depend on direction, fetch, water depth and other covariates (Mathiesen et al., 1994; Jonathan and Ewans, 2007;

Taylor et al., 2009; MacKay et al., 2010; Mazas and Hamm, 2011). Many possibilities exist for getting time series of homogeneous physical phenomena (clipping, decomposition, cleaning, etc.) but they are beyond the scope of this paper. In the following, if not stated differently, it will be considered that the time series are identically distributed, but they are still not composed of independent values.

Accordingly to the previous considerations and within the framework of the OTM, the constitution of the sample for the statistical inference of the extreme return levels of the environmental variable logically requires two successive and complementary steps: the physical declustering and the statistical optimization.

The need for both declustering and statistical optimization was generally recognized in the past. However they were often confused or merged together, arising methodological questions and confusing the meaning of the two operations. For instance, Lang et al. (1999) stated that “two different approaches can be adopted for threshold selection: the first one is based on physical criteria [...] and the second one is based on purely mathematical and physical considerations”, but they neither separate both steps, nor recognized their complementarity. It is indeed important to clarify which, among the numerous parameters to be defined for an OTM analysis, often largely arbitrary (Takvor and Panagiota, 2001), are involved in the physical declustering and which ones are involved in the optimization procedure. Even when in the past the two steps were performed separately (Dupuis, 1998; Egozcue et al., 2005; Bernardara et al., 2008, 2011; Garavaglia et al., 2010; Bardet et al., 2011; Mazas and Hamm, 2011; Wahl et al., 2011), or when approaching the two steps at the same time (i.e. the extremal index approach), the underlying concepts were not clearly exposed.

With the previous considerations in mind, this paper aims at clarifying the general conceptual framework of threshold selection for over-threshold modelling, distinguishing in particular the physical declustering procedure from the statistical optimization. A large literature review of the existing methods for both steps is given.

The main improvement of this effort of review and clarification is that, distinguishing both steps, the existing methods can be used in the right context, namely the physical declustering can be done based on physical arguments and the statistical optimization is performed later with purely statistical methods.

Note that this theoretical discussion is relevant in a multidisciplinary context, including different environmental applications (e.g. hydrology, meteorology, ocean sciences). This is an important point toward sharing of the knowledge of OTM techniques between different domains and different scientific communities. As a consequence, in this paper literature review and examples are based on environmental phenomena as different as floods, heavy rainfalls, extreme winds, high sea levels, extreme sea surges, and oceanic waves.

The paper is organized as follows. In Sect. 2, the two steps are depicted and the methodological framework is clarified. In Sect. 3 a review of the current methodology for physical declustering and statistical optimization of the statistical threshold is given. It is also shown that often in the past the two steps were merged together and sometimes confused. In Sect. 4 this general framework is applied to two different case studies for the estimation of hydrological and maritime extreme observations. In particular a double threshold is introduced. In Sect. 5 some final conclusions are drawn.

## 2 Distinguishing two steps for OTM

Let  $Z(t)$  be a time series (or, more generally, a spatial field) of discrete realizations of a given environmental variable,  $Z$  (e.g. a river discharge, a significant wave height, a sea level, a surge, a temperature, a wind speed, etc.) at a given resolution,  $\Delta t$  (i.e. time step, either regular or irregular). This can be the result of observation or modelling. The time series is assumed to be identically distributed.

For extreme value estimation applications, times series lasting several years are generally needed. For this reason, their size can be very important, depending on its duration,  $K$ , and time step,  $\Delta t$ . For example, a daily series lasting 20 yr contains around 7000 values, while an hourly series lasting 5 yr contains more than 40 000 values. Note also that environmental measures are often submitted to failure of measurement device, thus the time series can be incomplete.

### 2.1 Step 1: physical declustering

The physical declustering aims at extracting a sample of i.i.d. values,  $X_i$ , from the time series,  $Z(t)$ . This step can be viewed as an identification procedure, through purely physical consideration, of independent events. We claim that the notion of event and its correct understanding is a fundamental concept of EVT analysis applied to a time series. An event is defined here as a continuous physical phenomenon of the environmental variable, notably out of its mean regime, as can be instinctively comprehended by anyone: a storm for wind speed or wave height, a flood for river discharge, a heat or cold wave for temperature, a drought for rainfall, etc.

The events have a given duration that is often longer than the resolution  $\Delta t$  of the time series. In this case an event is composed of a set of consecutive discrete realizations of the variable, called cluster. The analyst should then define a random variable  $X$  describing the events. Very often this event-describing variable  $X$  represents the maximum value of  $Z(t)$  within the event, or cluster, and is often called the “peak” of the event. However,  $X$  may also be the result of any mathematical transformation of the cluster values: this is the case for the volume of a flood, for example. Generally speaking,  $X$  can be any characteristic of the event. The actual definition of

$X$  will depend mainly on the natural phenomenon involved, on the available data and on the aim of the study.

The appropriate physical declustering technique also depends on the characteristics of the given natural variable, of the time step of the series and on the physic and dynamic characteristics of the observed process. For instance, the declustering of a daily temperature time series will require different techniques than the declustering of hourly rainfall observations. In general, the knowledge of the physics of the studied phenomenon drives the physical declustering choices that should guarantee: first the independence of the selected events and second that no event is omitted in the process.

It is quite important to stress that  $Z$  and  $X$  are per se different random variables. It is quite clear in the case of daily river discharge vs. flood volume, but it is also true in the case of three-hourly wind speed vs. the peak wind speed of a storm, for instance. In particular, even if each event can be associated to a particular instant of occurrence on the time line,  $X$  does not depend on time any more.

A sample of  $N_T$  i.i.d. values  $X_i$  is thus obtained. Its size is much lower than the size of the time series  $Z(t)$ , generally in the order of few hundreds instead of several thousands.

In Sect. 3 a review of practical methods for physical declustering of environmental time series is given.

### 2.2 Step 2: statistical optimization

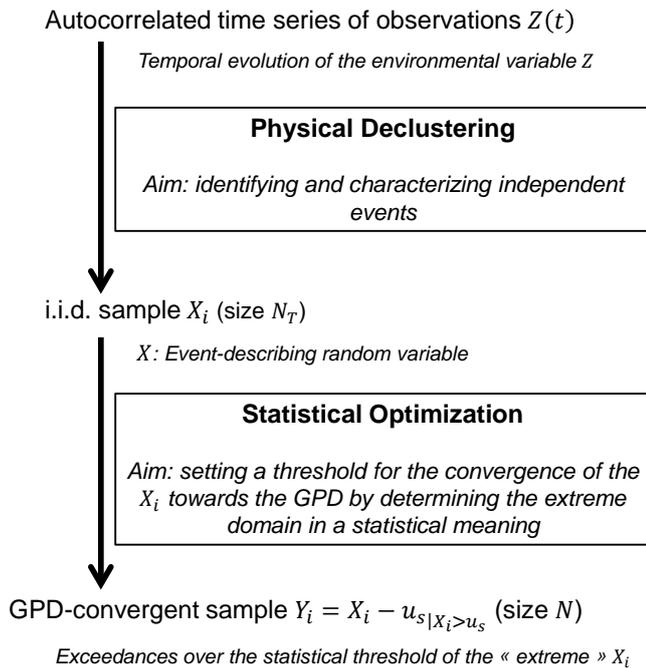
The physical declustering allowed the setting up of an i.i.d. sample: the EVT hypothesis is now met and the well-known statistical models can be applied.

Let us introduce  $u_s$ , which stands for statistical threshold and let us define the random variable  $Y = X - u_s$ , given  $X > u_s$ .  $Y$  is the exceedance of  $X$  above the threshold  $u_s$ . Thus a sample  $Y_i$  of size  $N$  can be defined from the sample  $X_i : Y_i = X_i - u_s$ , given  $X_i > u_s$ . Note the sample size reduction ( $N \leq N_T$ ) as the  $X_i$  values falling below (or equal to)  $u_s$  are excluded from the analysis.

Within the theoretical framework of OTM, and in particular according to Pickands’ theorem, when  $u_s$  increases, the probability distribution of the sample  $Y_i$  converges toward the generalized Pareto distribution (GPD) whose cumulative distribution function of the GPD, in its three parameter formulation, is given by

$$F(y) = 1 - \left[ 1 + k \left( \frac{y - \mu}{\sigma} \right) \right]^{-\frac{1}{k}}, \quad (1)$$

where  $k$ ,  $k \neq 0$ , is the shape parameter, also indicated as  $\xi$  (or  $-\xi$ ) in statistic literature,  $\sigma$  is the scale parameter and  $\mu$  is the location parameter, with  $y > \mu$  for  $k > 0$  and  $\mu < y < \mu - \sigma/k$  for  $k < 0$ . Note that following Pickand’s theorem, the location parameter is generally set equal to zero. Note also that the modified scale parameter,  $\sigma^* = \sigma - ku_s$ , is often used as scale indicator.  $u_s$  is thus the optimal threshold providing the best compromise between the convergence of the  $Y_i$  through a GPD (bias minimization) and the necessity



**Fig. 1.** General framework for identifying extreme data for over-threshold modelling (OTM).

to keep enough data  $Y_i$  for the estimation of its parameters (variance minimization). The extrapolation of the estimated GPD will yield the estimated return levels (or extreme quantiles). The  $u_s$  threshold selection step is called here statistical optimization. The statistical optimization step is a purely statistical problem for which several methods have been proposed in the literature; see Sect. 3 for a general review. It does not depend on the particular random variable (environmental or not) and it is general for every extreme value application.

A general overview on the two-step framework is depicted in Fig. 1.

### 3 Review of methods for physical declustering and for statistical optimization

In this section a literature review of the physical declustering and the statistical optimization techniques is given. Its aim is not to particularly recommend any of these to the detriment of the others, but rather to catalogue the different practices. At the end of this section a discussion is proposed to understand why and how these two steps were often merged and confused in the past and we point out that this knowledge is important in order to perform correctly both steps.

## 3.1 Methods for physical declustering (Step 1)

### 3.1.1 General principles

As stated in the previous sections, the physical declustering procedure aims at building the i.i.d. sample  $X_i$ , on which all the statistical analyses are based, by identifying and characterizing the events.

Generally speaking, this procedure does not require the “over-threshold” concepts per se. One could imagine, for instance, manually extracting a sample of extreme and independent events from a historical record of observations. Garavaglia et al. (2010) introduced the concept of central rainfall, defined as the rainfall observation  $Z(t)$  for which  $z(t-1) < z(t) > z(t+1)$  for declustering the rainfall observations series.

However, an overview of the literature (see in particular Lang et al., 1999) shows that the techniques based on the definition of a threshold to be exceeded spread widely. Adamowski (2000) claims, moreover, that the choice of the threshold for declustering is also a critical step in order to select homogenous events and to avoid merging different populations of observations (this is the concept of identical distribution mentioned above).

Following such an approach, the clusters (events) are usually defined as the series of consecutive values of  $Z(t)$  above a given threshold, which is called here  $u_p$ , for physical threshold. Note that this approach was often called partial duration series sampling in the past (Cunnane, 1973; Rosbjerg, 1985; Rosbjerg et al., 1992; Madsen and Rosbjerg, 1997).

As mentioned in Sect. 2.1, the event-describing variable  $X$  could be any transformation of the values of  $Z$  within the cluster. For example, a temporal integration of the consecutive value of river discharge over the physical threshold is sometimes used in hydrology to characterize the volume of a flood. However, in most environmental applications, the maximum value of  $Z(t)$  observed during the cluster  $i$ , or event peak  $X_i$ , is retained to describe the event. For this reason the name of peak over threshold (POT) spread widely in the literature for this sampling technique.

It has been stressed above that the physical declustering must not only identify the events, but also guarantee their independence. As a consequence, the actual definition of a cluster (event) generally relies upon two different families of parameters: on the one hand, the physical threshold  $u_p$ , whose value is expressed in the same units as  $Z(t)$  (but not necessarily as  $X_i$ , e.g. the volume of a flood); on the other hand, one or several parameters needed for ensuring the independence of the different clusters.

Several physically based criteria are available in the literature for the definition of the physical threshold and for the characterization of the independence of the clusters.

### 3.1.2 Physical threshold

The choice of the threshold  $u_p$  is a key step for the identification of events. Several approaches were proposed in the past for the definition of the physical threshold.

First, and quite simply, the choice of the physical threshold can be defined by an expert prior knowledge. Expert prior knowledge can be based, in a heuristic approach, on the practical consequences for the phenomenon to cross a given threshold value. For example, Bonazzi et al. (2012) define a wind velocity threshold corresponding to the level at which damage to buildings is likely to occur. Perception threshold, defined by expert estimation, is used to perform OTM for historical and non-systematic observation, (Ouarda et al., 1998; Barriendos et al., 2003; Payrastre et al., 2005, 2011; Hamdi et al., 2013).

For threshold selection, Lang et al. (1999) defined the mean number of events per year,  $\lambda_T = N_T/K$ , where  $K$  is the total duration of the time series, in years, and  $N_T$  is the number of physical events to be selected (and also the sample size of the  $X_i$ ). Then they discussed the evolution of  $\lambda_T$  when the threshold increases and they identified four domains. First, the threshold is below the minimum of the time series: thus the entire time series is considered as an event, though it has no physical sense. Second, as the threshold value raises between the minimum and (roughly) the mean value of the series  $\lambda_T$  increases: the higher the threshold, the more events there are. This actually means that just shortfalls below a low threshold are identified. Third,  $\lambda_T$  reaches a maximum (when the threshold is close to the mean of the series) and begins to decrease. Fourth, when the threshold is larger than the maximum of the series, no more exceedances can be extracted and  $\lambda_T = 0$ . The authors require that  $\lambda_T$  be in the third domain, though far from both the lower and upper limits of the domain. In the proposed conceptual framework, it can be stated that events (as defined above) are identified in the third domain. However, though this recommendation is relevant, it is easily fulfilled and is not specific enough to be really useful in practical applications.

Another widely used approach relies on the idea that  $u_p$  can be tuned in order to obtain a physically reasonable value of  $\lambda_T$ . More generally, the choice of the actual number  $\lambda_T$  is based on expert knowledge of the physics and the dynamics of the process. For example, in hydrological applications it is suggested to choose  $\lambda_T < 5$ , which is a large number of floods to observe, for a given site, on average, per year. Obviously, this number should depend on an on-site hydrological regime. Working on skew surges, in a regional analysis framework, (Bernardara et al., 2011) suggested to set  $u_p$  so that  $\lambda_T = 1$ , while for local analysis Walton (2000) fixed  $\lambda_T = 3$ . Analyzing significant wave heights, (Mazas and Hamm, 2011) suggested that the value of  $\lambda_T$  should be roughly between 5 and 10, also depending on the value of the time series duration (closer to 5 for long time series, closer to 10 for short ones). Tawn and Vassie (1989) suggested a value

of  $\lambda_T$  around 5 for extreme sea surge. Floris et al. (2010) analyzed  $\lambda_T$  in a framework of extreme rainfall analysis.

Some authors set the threshold using a given quantile of the time series  $Z(t)$ . Ruggiero et al. (2010) set it to the 99.5th percentile of the data, working with wave heights. Rosbjerg et al. (1992) suggest calculating the physical threshold as the mean value of the observed series plus three standard deviations. Notice that, though it looks like a statistical approach, there is no optimization process in it.

These criteria for the selection of the relevant physical threshold were compared and simultaneously used in the past, for instance Ntegeka and Willems (2008) stated that “an extreme event can be selected based on frequency, intensity, threshold exceedances or physical expected impacts”.

### 3.1.3 Parameters for ensuring the independence of the events

In order to ensure the independence of the selected events, many physically based criteria have been developed. Several of these criteria are recurrent in the literature and will be presented here.

The most common techniques consist in setting temporal parameters, most of the time based on the minimal time lag between two events. The idea is quite simple: after a given period of time, the autocorrelation between the observations becomes negligible and two events can be safely considered independent. The definition of this time lag is directly derived from the physics of the natural phenomenon: it should be longer than the typical duration of the physical processes (usually meteorological ones) generating the events. Thus it can be set by an expert prior knowledge. However, the time lag should not be too long in order to avoid discarding independent events and thus missing valuable information. For instance, in north-eastern Europe, extreme wave heights may be generated by successive storms moving along the storm track every 24 h or so; therefore setting the time lag to 48–72 h could lead to miss information. Many applications of this approach are available in the literature: Egozcue et al. (2005) studied wave height hazards along the Mediterranean coast of Spain and set the time lag to 4 days; Haigh et al. (2010) studied the extreme sea levels along the English Channel and required the surge peaks to be separated by 30 h at least; USWRC (1976), Cunnane (1979) and Lang et al. (1999) imposes that successive river flood events be separated by at least as many days as five plus the natural logarithm of square miles of the basin area. Willems (2000) required that two rainfall events are separated by at least a 12 h lag.

The time lag can also be defined using the autocorrelation function of the time series  $Z(t)$ . For extreme wave heights, Mathiesen et al. (1994), propose requiring that it cannot be longer than the time interval for which the autocorrelation function of the series drops under 0.3–0.5. In a similar way, while studying storm surge extremes along the US East

Coast, Walton (2000) established the typical duration of an event via the autocorrelation of the surge series and found that the drop off of the autocorrelation function to a noise level value close to zero was on the order of 24–72 h. Note that some authors define the time lag between the end of an event and the beginning of another while others define it between two peaks.

In general, both physical and statistical methods aim at the definition of the correlation length, thus they should naturally converge.

Other temporal parameters can be used; for instance, in their atlas of waves along the Italian coasts, (Franco et al., 2004) allowed short fluctuations of the time series below the threshold  $u_p$  up to a maximal duration (6 h) and also set a minimal storm duration (12 h). This last parameter may be useful for some applications; for instance, waves generated by a very short storm will not cause damage to a breakwater. In the sea wave analysis field, (Takvor and Panagiota, 2001) extracted independent the sea state by looking at wave energy reductions between consecutive time steps. In contrast, (Smith, 1988) examined the typical duration of extreme wave conditions and did not see any rationale for using such a parameter.

Another technique consists in using a secondary threshold: in this approach, two events are considered independent when the signal  $Z(t)$  falls below this value. In particular, this secondary threshold may be defined as a fraction of the physical threshold  $u_p$  (in this case the value of this fraction can indeed be considered as the parameter to set) or as a fraction of the peak value of one of these events. For instance, (USWRC, 1976) requires (among other criteria) that the intermediate flows between two consecutive flood peaks must drop below 75 % of the lowest of these two peaks, while (Cunnane, 1979) imposes that the flow must drop below 2/3 of the first peak value.

The independence of the selected events (or more generally the independence of the selected clusters) has been checked in the past via the analysis of the probability distribution of the occurrences of the events for a given time interval. In fact, the Poisson distribution (Haight, 1967) is a discrete probability distribution that expresses the probability of a given number of events occurring in a fixed interval of time if these events occur with a known average rate and independently. Thus, if the number of occurrences of events follows a Poisson distribution, the events are supposed to be independent (Cunnane, 1979; Rosbjerg et al., 1992; Lang et al., 1999). Some authors, (e.g. Ashkar and Ouarda, 1996, Silva et al., 2011) selected the physical threshold corresponding to the best adaptation of the number of occurrences to the Poisson distribution and checked the independence hypothesis looking at the uniform distribution of the arrival time of events depending on the observation period support.

Note that for several of these parameters, their value should be somehow dependent on the value of  $u_p$ . For instance, if the time lag between two events is defined between

the end of the first and the beginning of the second one, or if a fraction of  $u_p$  is used to ensure the independency, the values to be considered could be different if  $u_p$  is rather low or rather high. This is particularly true for the minimal event duration or the maximal duration of fluctuations below  $u_p$ , even though these parameters are quite scarcely used.

### 3.2 Methods for statistical optimization (Step 2)

Once the sample of the i.i.d. data  $X_i$  is built, the statistical optimization consists in choosing the relevant value of the statistical threshold  $u_s$  to take into account for the estimation of the GPD model on the observations exceeding the threshold, toward which the sample is supposed to converge (Beirlant et al., 1996).

The threshold selection criteria here are statistically based and they aim at meeting the EVT hypothesis and the best compromise between bias and variance. In this step, the question is which ones of these events are extreme from a statistical point of view?

A first class of such methods is based on the maximization of the goodness of fit between the probability distribution and the data or the minimization of the asymptotic mean square error of the estimators. Several authors suggested choosing the value of  $u_s$  providing the best GPD adaptation to the empirical data. That can be done through the optimization of  $\chi^2$  or the Kolmogorov–Smirnov test. For instance, (Bernardara et al., 2011) used this approach for fitting a regional surge probability distribution. Anderson–Darling (AD) goodness-of-fit test is suggested and employed by Choulakian and Stephens (2001) and Haylock (2011). The adaptation of GPD to empirical data above the threshold could also be checked via the L-moments. In particular, for GPD the relation between L-moments of order 3 and 4 is known and L-moments plot technique can be used (Hosking and Wallis, 1997).

A similar class of methods are based on the minimization of the variance estimation of the Hill semi-parametric estimator of the tail index, (Hill, 1975; Hall, 1982; De Haan and Peng, 1998). (Beirlant et al., 1996) and (Willems 1998) introduced a systematic methodology based on these principles. In (Willems, 2000) an application to rainfall observations is given. (Neves and Fraga Alves, 2004) give a short review on these methods and they propose an automatic selection procedure.

A well-known property of the GPD is that the shape and modified scale parameters will remain constant when the threshold increases. Following this property, (Davison and Smith, 1990; Lang et al., 1999; Egozcue et al., 2005) suggested choosing the threshold so that the mean of the exceedances above the threshold,  $E(X - u_s)$ , is a linear function of the threshold value, indicating a range where the GPD parameters are not depending on the threshold selection. This technique is also known as MRL (Mean Residual Life) plot, (Coles, 2001). In this framework, (Begu eria, 2005) chose the

value of  $u_s$  in maximizing the fit of a linear function on the mean excess function.

(Coles, 2001) also proposed the STM (stability method) consisting in defining an optimum where the shape parameter of the GPD distribution is approximately constant for small threshold changes. (Mazas and Hamm, 2011) performed a sensitivity analysis of shape and modified scale parameters with respect to the (statistical) threshold in order to identify domains of stability. Then they selected the lowest threshold (minimization of variance) of the highest domain of stability (minimization of bias). (Thompson et al., 2009) used the property of stability of the GPD modified scale parameter to introduce a procedure for automatizing the threshold selection. They define 100 equally spaced threshold values between the median of the time series  $Z(t)$  and its 98th percentile. For each value, the stability of the modified scale parameter is tested by the Pearson normality tests. The first (i.e. lowest) value satisfying the test is retained as the statistical threshold.

The stability of some relevant quantiles of the GPD distribution has been used in the past for the selection of the optimal  $u_s$  value, for instance by Rosbjerg et al. (1992), among many others.

### 3.3 Review discussion

The concept of the exceedances over a given threshold was used both for physical declustering and statistical optimization (Smith, 1984; Lang et al., 1999; Parent and Bernier, 2003).

This is explained by the fact that this concept is, on the one hand, useful for defining events as independent clusters of observations and, on the other hand, is also consistent with the EVT concepts of GPD convergence.

However, the use of the same concept of “exceedances over a threshold” for the two different steps of the analyses led to some incoherencies and confusions.

First of all, as pointed out in the previous section, it should be highlighted that the domains of application of the two steps are different. The declustering procedure applies to highly correlated data, such as the environmental time series of observations and it was studied mainly by earth scientists in the past. The statistical threshold optimization applies to a large number of statistical problems as it was treated in the past mainly by the statistical community. That leads to some incoherencies on the vocabulary used and on the concepts definition. For example, (Takvor and Panagiota, 2001), in a review of declustering techniques, called the physical declustering “statistical pre-processing”, a definition that can confuse the reader. In a similar way, the concept of “POT method” often includes the whole methodology for the determination of the probability of occurrence of the extremes values, including the GPD model, while it should be restricted to the declustering step. The same comment holds for the “partial duration series” definition which refers to the first part

of physical declustering but it was used in the past to indicate the whole analysis. Also, note that following the EVT vocabulary, the word “extreme” is restricted to the values exceeding the threshold  $u_s$  while the full sample of the  $X_i$  represent the whole i.i.d. population describing the different events. However, in practice, the  $X_i$  are often arbitrarily considered as an “extreme” population, which may be confusing.

Another example of incoherencies can be found in the number of data to be selected. As explained in section 3.1, (Lang et al., 1999) pointed out that the number of peaks over the threshold,  $X_i$  can decrease but also increase when the threshold  $u_p$  increases. This is logical in a physical point of view, because the sample of the  $X_i$  completely changes depending on the value of  $u_p$ . However, for the statistical optimization threshold  $u_s$  the number of exceedances over the threshold  $Y_i$  must decrease when the threshold  $u_s$  increases. Introducing the concept of physical event to be identified in the time series is thus much relevant for understanding such a behaviour.

Note also that, following the statistical theory (Pickands, 1975), the EVT requires choosing all the values over a given threshold,  $u_s$ , and not only some of them, as in the case of physical declustering using the peaks over  $u_p$ .

Moreover, the EVT states that the  $Y_i$  sample converges towards a GPD distribution, while the  $X_i$  sample, representing a whole population of i.i.d. events could be described by any statistical model. Indeed, several authors (Goda, 1988, 2010, 2011; Mathiesen et al., 1994; Goda et al., 2010) considered other distributions than the GPD (or GPD family).

In practice, another strong rationale arises for separating both steps. As shown by the literature review in Sect. 3.2, most of the methods for determining the statistical threshold require testing many values of  $u_s$ . If the physical declustering has not been performed prior to this, it will have to be done as many times as there are tested values of  $u_s$ , instead of just once. If one keeps in mind that the declustering of time series, whose size can be up to several hundreds of thousands of data, can be quite computer intensive, the interest of running this step once for all instead of 10 to 100 times is obvious. Furthermore, it has been shown in Sect. 3.1 that during the declustering process, the parameters ensuring the independence of the events may depend on the  $u_p$  threshold value, or be relevant for a small range of  $u_p$  values only. Then there is much interest in setting them accurately with regard to the value of  $u_p$  once and for all, instead of repeating this many times. This will be illustrated in the second case study in Sect. 4.3.

Note that the clusters population, namely the structure of the events, described by the clusters of  $Z(t)$  over  $u_p$ , present an interest in itself. It is important, for some application to calculate some statistics such as the probability distribution of the event size or duration, the internal correlation or the shape characteristics of a general event. This shows again the interest in separating the two steps.

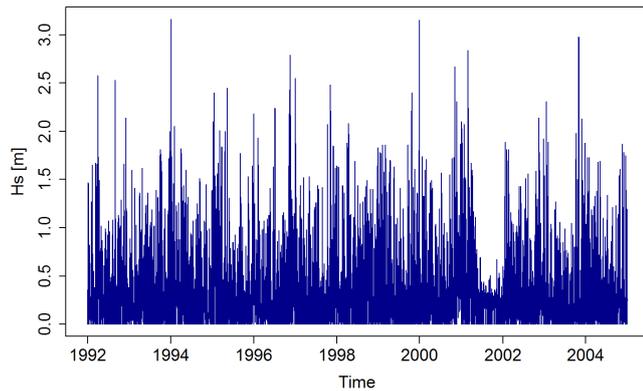


Fig. 2. Swell significant wave height time series off Marseille.

Even though the methodological framework was not yet clarified, in the past few authors highlighted some points which direct attention to the different meaning and application of the two steps. A very good illustration is provided by Cruise and Arora (1990) who noticed that the threshold level for physical declustering often had to be raised significantly to meet exponentially based tests on the POT distribution (Lang et al., 1999).

Some other authors in the past proposed extreme value analyses in which the two steps are well distinguished. For example (Dupuis, 1998; Egozcue et al., 2005; Bardet et al., 2011; Bernardara et al., 2011; Mazas and Hamm, 2011) applied two different thresholds, one for the physical declustering, the other for the statistical optimization. (Garavaglia et al., 2010) used the central rainfall concept for physical declustering and arbitrarily fixed the  $u_s$  at the 70 % quantile of the distribution, without any optimization. (Bernardara et al., 2008) used classical declustering criteria for daily discharge series and a specific optimization for the shape parameter of the GPD distribution.

However, the theoretical framework distinguishing two steps was not clearly defined. Hence, as a conclusion of this review effort, we deem that the lack of the concept of event is a major cause of the confusion observed in the past. It is useful for the understanding of EVT analyses of auto-correlated time series and it has a sound physical basis.

## 4 Applications to environmental variables

### 4.1 The double threshold approach

In order to illustrate the proposed framework, and in coherence with the literature, we propose here to provide both physical declustering and statistical optimization through a threshold approach.

This approach is applied to two case studies, a wave height study and an extreme discharge study, in order to illustrate

Table 1. Summary of parameters for the two case studies.

	Wave heights, Marseille	Discharge/flood volume, Rieutord
$K$ [yr]	13	19.33
$\Delta t$	3 h	1 day
$n$	38 005	7062
$u_p$	1.4 m	$10 \text{ m}^3 \text{ s}^{-1}$
$\lambda$ [events $\text{yr}^{-1}$ ]	10	3.6
$N_T$	130	70
$u_s$	1.87 m	$6.48 \text{ Mm}^3$
$N$	43	25

that the methodological framework is valid in different fields of natural hazard estimation.

It is shown that this approach allows selecting the correct techniques and carrying out a complete analysis, extracting all the relevant information.

### 4.2 Wave heights

We consider in this first illustrative example a time series of simulated three-hourly significant wave heights  $H_s$  offshore Marseille, France ( $5.3104^\circ \text{ W}$ ,  $43.3460^\circ \text{ N}$ ; water depth: 34 m). The duration of the data is  $K = 13$  yr, and the size of the time series is  $n = 38,005$  data. In order to ensure the homogeneity of the data, a decomposition of the sea states have been performed and only the swell component have been retained. The  $H_s$  time series is plotted in Fig. 2.

In this case study,  $Z$  is a three-hourly significant wave height of the swell component (in metres), the events to be identified are swell storms and they are classically described by the random variable  $X$  “storm peak” (in metres), that is, the maximum value within the cluster. A physical threshold has been set in order to obtain a sample of  $\lambda_T = 10$  storms per year in average, which is a physically sounding number of extreme events per year for the region. The physical threshold is thus set to  $u_p = 1.4$  m. The declustering has been performed by using a minimal duration of 24 h between two storms to ensure their independence. Furthermore, a minimal storm duration of 6 h has been set (because very short events do not cause important damage to coastal structures) and fluctuations below the threshold within a same storm have been allowed for less than 12 h. These parameters can be considered relevant for the chosen physical threshold but it would not be the case for a higher or lower threshold. It is important to notice that this first step allowed defining a population of events which can be analyzed to extract relevant statistics. In particular, the extraction yields  $N_T = 130$  events, the mean storm duration of the events is around 15 h. In all, 96 % of the events last less than 36 h, which is consistent with the physics. Unsurprisingly, a strong seasonality is observed: 28 storms in fall, 38 in winter, 23 in spring and 4 in summer.

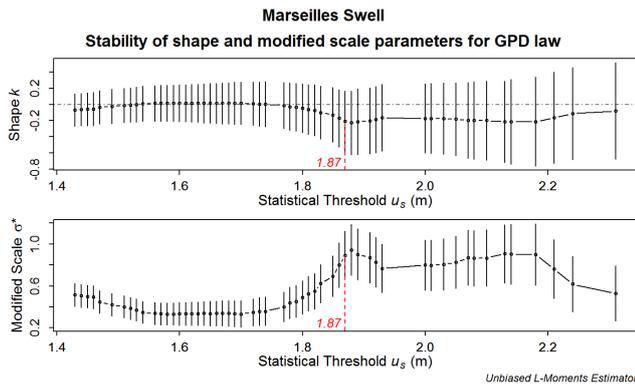


Fig. 3. Stability of shape parameter and modified scale parameter for Marseille series of swell waves.

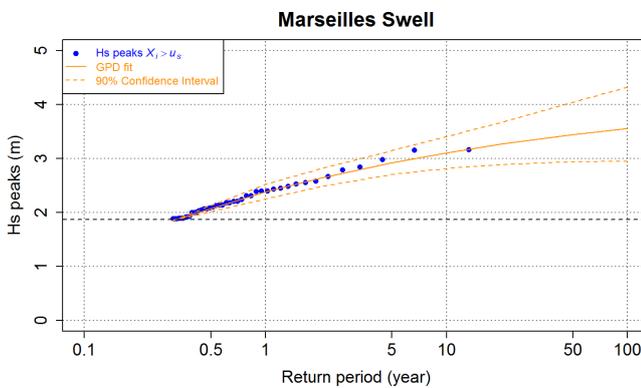


Fig. 4. GPD fit for the swell off Marseille.

The statistical optimization is then performed on the sample  $X_i$ . Note that in this case,  $Z$  and  $X$  have the same dimension; since  $X_i \geq u_p$  it can thus be written that  $u_p \leq u_s$  (meaning that testing values of  $u_s < u_p$  would be useless). The stability of the GPD shape parameter  $k$  and modified scale parameter  $\sigma^* = \sigma - ku$  with respect to the statistical threshold  $u_s$  is illustrated in Fig. 3, along with the associated 95 % confidence intervals computed by the asymptotic method. A first “domain of stability” can be seen between roughly 1.5 and 1.8 m, then a second one between 1.87 and 2.2 m. Afterwards the sample size is too short and the parameter uncertainty is too great. The bias minimization requires to choose the highest domain of stability while the variance minimization needs as much data as possible; consequently, the statistical threshold is set to  $u_s = 1.87$  m, yielding  $N = 43$ . Note that, here  $u_s = u_p + 0.47$  m and the number of observation has been reduced from  $n = 38,005$  to  $N = 43$ .

The GPD parameters are estimated by the L-moments estimator (Hosking and Wallis, 1997). The fit is illustrated in Fig. 4. A summary of the parameters estimate for this case study is given in Table 1.

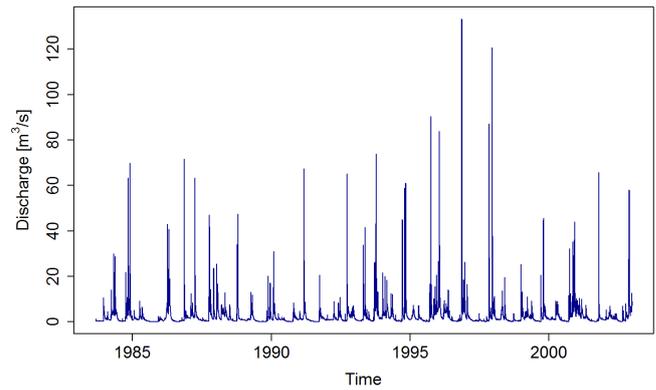


Fig. 5. Discharge series of the Loire river at Rieutord.

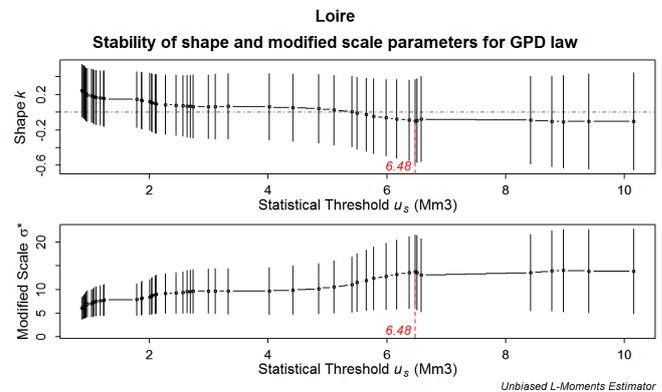


Fig. 6. Stability of shape and modified scale parameters of Loire river sample.

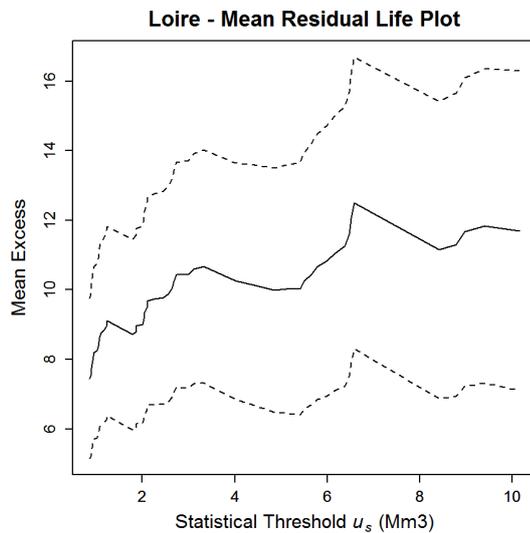
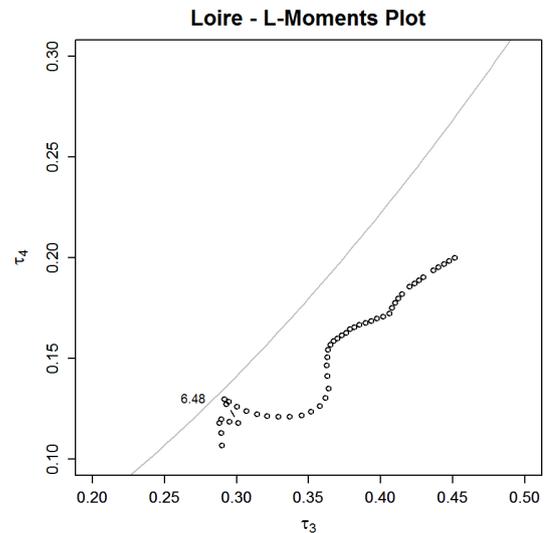
### 4.3 Discharge and flood volumes

The selected time series is the Loire river discharge daily series at Rieutord, from 01/09/1983 to 31/12/2002. The duration of the data is  $K = 19.33$  yr, and the size of the time series is  $n = 7062$  data. The discharge series is plotted in Fig. 5. The local modulus of the river is estimated around  $2.7 \text{ m}^3 \text{ s}^{-1}$ . A physical threshold of  $u_p = 10 \text{ m}^3/\text{s}$  based on an expert judgment has been applied. The inter-event duration has been set to 10 days. That leads to an average number of flood events per year  $\lambda_T = 3.6$ , which is physically acceptable and coherent with expert prior knowledge based on physical characteristics of the discharge phenomenon.  $N_T = 70$  flood events are thus retained. Their mean duration is around 103 h. 50 % of the flood peaks last only one day, while 9 % last two days and 14 % last three days. A strong seasonality can be observed, with 27 flood peaks in fall, 23 in winter, 19 in spring and only one in summer.

It is decided to describe these flood events by their volume:  $X$  is the temporal integration of the discharge over the flood duration. In particular, the physical threshold is assumed to define completely the flood, allowing the computation of its volume. In this case study,  $Z$  is a daily discharge in  $\text{m}^3 \text{ s}^{-1}$ ,

**Table 2.** Evolution of the size of the upper part of the  $X_i$  sample with respect to  $u_p$ .

$u_p$	Number of events whose peak is above				
	$5 \text{ m}^3 \text{ s}^{-1}$	$10 \text{ m}^3 \text{ s}^{-1}$	$15 \text{ m}^3 \text{ s}^{-1}$	$20 \text{ m}^3 \text{ s}^{-1}$	$25 \text{ m}^3 \text{ s}^{-1}$
$5 \text{ m}^3 \text{ s}^{-1}$	75	52	37	33	26
$10 \text{ m}^3 \text{ s}^{-1}$	–	70	47	40	32
$15 \text{ m}^3 \text{ s}^{-1}$	–	–	49	41	33
$20 \text{ m}^3 \text{ s}^{-1}$	–	–	–	42	34
$25 \text{ m}^3 \text{ s}^{-1}$	–	–	–	–	34

**Fig. 7.** Mean residual life plot for the flood volumes of Loire river sample.**Fig. 8.** L-moments plot for the flood volumes of the Loire river sample.

an event is a flood and  $X$  is the volume of the flood, in  $\text{Mm}^3$ . This is of course a very rough estimate of the flood volume: in a genuine hydrological study, it should be computed based on the hydrogram of each flood. This application is for illustrative purpose.

The stability of GPD parameters with respect to the statistical threshold  $u_s$  is given in Fig. 6. The shape (resp. modified scale) parameter slowly decreases (resp. increases) from about  $1 \text{ Mm}^3$  to  $6.48 \text{ Mm}^3$ , then remains remarkably constant. This result is confirmed by MRL Plot, Fig. 7.  $L$  moment analysis depending on the statistical threshold  $u_s$  is shown in Fig. 8. Here it is found that the threshold value of  $6.48 \text{ Mm}^3$  is the only one for which the  $L$  moments are almost lying on the theoretical GPD curve. Thus the statistical threshold is set to  $u_s = 6.48 \text{ Mm}^3$ , yielding a sample of  $N = 25$  flood volume exceedances  $Y_i$ . In Fig. 9, the corresponding GPD calibration is shown, along with the 90% confidence interval. A summary of the parameters estimate for this case study is given in Table 1.

In this case study, the inter-event duration is defined between the end of an event and the beginning of another. As

has been stressed in Sect. 3.1.3, a consequence is that not only the lower part but also the upper part of the  $X_i$  sample, and thus ultimately of the  $Y_i$  sample to be fit, may vary when a broad range of  $u_p$  values is tested while the independence criteria remain constant. This is the case for this declustering procedure on this sample. Different values of  $u_p$  have been tested: 5, 10, 15, 20 and  $25 \text{ m}^3 \text{ s}^{-1}$ . For each value, a number of events are identified. Then, for each of these  $X_i$  samples (the events are here described by the discharge peaks), the number of events exceeding a higher value than the threshold is counted. The results are given in the Table 2. For instance, if the physical threshold is set to  $5 \text{ m}^3 \text{ s}^{-1}$ , the number of peaks exceeding  $25 \text{ m}^3 \text{ s}^{-1}$  is 26, while, if the physical threshold is set to  $25 \text{ m}^3 \text{ s}^{-1}$ , their number increase to 34. This illustrates that one should be careful when choosing or tuning the independence criteria and this is an additional reason for separating the physical declustering step from the statistical optimization step, all the more since the latter one often requires investigating a wide range of threshold values.

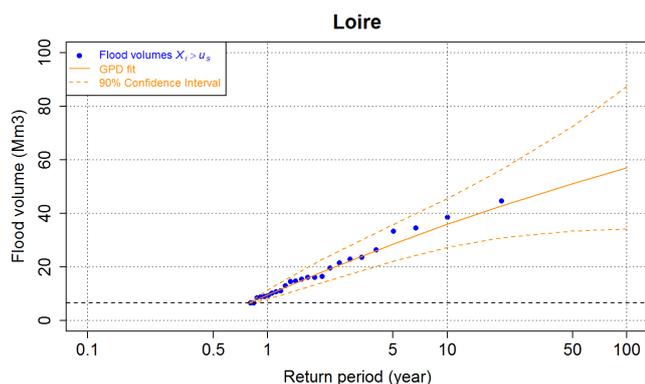


Fig. 9. GPD fit for the flood volumes of the Loire river at Rieutord.

## 5 Conclusions

This paper clarifies the general framework for “over the threshold” exceedance modelling, distinguishing in particular the physical declustering procedure to the statistical optimization. The two steps have a very different meaning and very different techniques have been proposed in the past for the application of these two steps. The literature is wide and an effort of review of the existing methods for both steps is done. It allows choosing the most appropriate methods for each step. In particular, we highlighted that declustering techniques are mostly based on the analysis and on the characterization of the physics of the phenomenon, while statistical optimization is a purely statistical problem. A consequence is the importance of the notion of physical event. It was often underlying in the literature, but we deem it most important to make it explicit. From our point of view, the distinction between the auto-correlated observations, at a regular time step, of the time series and the independent and self-consistent physical events should become central in the extreme value analysis of environmental variables. We also claim that fully apprehending both the difference and complementarity of these two steps allows a clearer understanding of the meaning of the different available tools and of the parameters needed for the rest of the analysis (i.e. the independence criteria parameters) and thus, ultimately, the appropriate application of the existing threshold selection methods. Through two simple practical examples, we show that each step has a clearly distinct role. It is also worthy to note that this discussion is relevant for several domains of natural hazard estimation, and even more generally to any EVT analysis of auto-correlated time series.

Edited by: T. Glade

Reviewed by: P. Jonathan and one anonymous referee

## References

- Adamowski, K.: Regional analysis of maximum and partial duration flood data by nonparametric and L-moments methods, *J. Hydrol.*, 229, 216–223, 2000.
- Allamano, P., Laio, F., and Claps, P.: Effects of disregarding seasonality on the distribution of hydrological extremes, *Hydrol. Earth Syst. Sci.*, 15, 3207–3215, 2011.
- Ancona-Navarrete, M. A. and Tawn, J. A.: A comparison of methods for estimating the extremal index, *Extremes*, 3, 5–38, 2000.
- Ashkar, F. and Ouarda, T. B. M. J.: On some methods of fitting the generalized Pareto distribution, *J. Hydrol.*, 177, 117–141, 1996.
- Bardet, L., Duluc, C. M., Rebour, V. and L’Her, J.: Regional frequency analysis of extreme storm surges along the French coast, *Nat. Hazards Earth Syst. Sci.*, 11, 1627–1639, 2011.
- Barriados, M., Coeur, D., Lang, G., Llasat, M. C., Naulet, R., Lemaitre, F., and Barrera, A.: Stationarity analysis of historical flood series in France and Spain (14th–20th centuries), *Nat. Hazards Earth Syst. Sci.*, 3, 583–592, 2003.
- Beguéria, S.: Uncertainties in partial duration series modelling of extremes related to the choice of the threshold value, *J. Hydrol.*, 303, 215–230, 2005.
- Beirlant, J., Goegebeur, Y., Segers, J., and Teugels, J. L.: *Statistics of Extremes, Theory and Applications*, 2004.
- Beirlant, J., Teugels, J. L., and Vynckie, P.: *Practical analysis of Extreme Values*, Leuven University Press, 1996.
- Bernardara, P., Andreewsky, M., and Benoit, M.: Application of the Regional Frequency Analysis to the estimation of extreme storm surges, *J. Geophys. Res.*, 116, C02008, doi:10.1029/2010JC006229, 2011.
- Bernardara, P., De Michele, C., and Rosso, R.: A simple model of rain in time: An alternating renewal process of wet and dry states with a fractional (non-gaussian) rain intensity, *Atm. Res.*, 2006.
- Bernardara, P., Schertzer, D., Sauquet, E., Tchiguirinskaia, I., and Lang, M.: The Flood probability distribution tail: how heavy is it?, *Stoch. Environ. Res. Risk Assess.*, 22, 107–122, 2008.
- Bonazzi, A., Cusack, S., Mitas, C., and Jewson, S.: The spatial structure of European wind storms as characterized by bivariate extreme-value Copulas, *Nat. Hazards Earth Syst. Sci.*, 12, 1769–1782, 2012.
- Choulakian, V. and Stephens, M. A.: Goodness-of-fit tests for the Generalized Pareto Distribution, *Technometrics*, 43, 478–484, 2001.
- Coles, S.: *An introduction to statistical modeling of extreme values*, London, Springer-Verlag, 2001.
- Colombo, A. F., Etkin, D., and Karney, B. W.: Climate Variability and the Frequency of Extreme Temperature Events for Nine Sites across Canada: Implications for Power Usage, *J. Climate*, 12, 2490–2502, 1999.
- Cruise, J. F. and Arora, K.: A hydroclimatic application strategy for the Poisson partial duration model, *Water Resour. Bull.*, 26, 431–442, 1990.
- Cunnane, C.: A particular comparison of annual maxima and partial duration series methods of flood frequency prediction, *J. Hydrol.*, 18, 257–271, 1973.
- Cunnane, C.: A note on the Poisson assumption in partial duration series models, *Water Resour. Res.*, 15, 489–494, 1979.
- Davison, A. C. and Smith, R. L.: Models for exceedances over high thresholds, *J. R. Stat. Soc. B*, 52, 393–442, 1990.

- De Haan, L. and Peng, L.: Comparison of tail index estimators, *Statist. Neerlandica*, 52, 60–70, 1998.
- Dupuis, D. J.: Exceedances over High Thresholds: A Guide to Threshold Selection, *Extremes*, 1, 251–261, 1998.
- Egozcue, J. J., Pawlowsky-Glahn, V., and Ortego, M. I.: Wave-height hazard analysis in Eastern Coast of Spain – Bayesian approach using generalized Pareto distribution, *Adv. Geosci.*, 2, 25–30, 2005.
- Elek, P. and Markus, L.: A long range dependent model with non-linear innovations for simulating daily river flows., *Nat. Hazards Earth Syst. Sci.*, 4, 277–283, 2004.
- Embrechts, P., Klüppelberg, C., and Mikosch, T.: Modelling extremal events, *Stochastic Modelling and Applied Probability*, Springer-Verlag, 33, p 648, 1997.
- Floris, M., D'Alpaos, A., Squarzone, C., Genevois, R., and Marani, M.: Recent changes in rainfall characteristics and their influence on thresholds for debris flow triggering in the Dolomitic area of Cortina d'Ampezzo, north-eastern Italian Alps, *Nat. Hazards Earth Syst. Sci.*, 10, 571–280, 2010.
- Franco, L., Piscopia, R., Corsini, S., and Inghilesi, R.: L'Atlante delle onde nei mari italiani - Italian Wave Atlas, Full Final Report by APAT-University of Roma Tre, sponsored by AIPCN Italian Section and Italia Navigando, 2004.
- Fréchet, M.: Sur la lois de probabilité de l'écart maximum, *Annales de la société polonaise de mathématique (in French)*, 6, 93–122, 1928.
- Garavaglia, F., Gailhard, J., Paquet, E., Lang, M., Garçon, R., and Bernardara, P.: Introducing a rainfall compound distribution model based on weather pattern sub-sampling, *Hydrol. Earth Syst. Sci.*, 14, 951–964, 2010.
- Gnedenko, B.: Sur la distribution limite du terme maximum d'une série aléatoire (in French), *Ann. Math.*, 44, 423–453, 1943.
- Goda, Y.: On the Methodology of Selecting Design Wave Height, *Proc. 21st Int. Conf. on Coast. Eng.*, Malaga, Spain, 1988.
- Goda, Y.: Random seas and design of maritime structures, Singapore, World Scientific Publishing, 2010.
- Goda, Y.: Plotting-position estimator for the L-moment method and quantile confidence interval for the GEV, GPA, and Weibull distributions applied for extreme wave analysis, *Coast. Eng. J.*, 53, 111–149, 2011.
- Goda, Y., Kudaka, M., and Kawai, H.: Incorporating of Weibull distribution in L-moments method for Regional Frequency Analysis of Peak Over Threshold wave heights, ICCE, Shanghai, 2010.
- Gumbel, E. J.: *Statistics of Extremes*, New York, Columbia University Press, Columbia, 1958.
- Haigh, I. D., Nicholls, R., and Wells, N.: A comparison of the main methods for estimating probabilities of extreme still water levels, *Coast. Eng.*, 57, 838–849, 2010.
- Haight, F. A.: *Handbook of the Poisson Distribution*, New York, John Wiley & Sons, 1967.
- Hall, P.: On some simple estimate of an exponent of regular variation, *J. R. Stat. Soc.*, 44, 37–42, 1982.
- Hamdi, Y., Duluc, C.-M., Deville, Y., Bardet, L., and Rebour, V.: Use of historical information in extreme storm surges frequency analysis, EGU General Assembly, Vienna, Geophysical Research Abstracts, 2013.
- Haylock, M. R.: European extra-tropical storm damage risk from a multi-model ensemble of dynamically-downscaled global climate models, *Nat. Hazards Earth Syst. Sci.*, 11, 2847–2857, 2011.
- Hill, B. M.: A simple General Approach to Inference About the Tail of a Distribution, *The Annals of Statistics*, 3, 1163–1174, 1975.
- Hosking, J. R. M. and Wallis, J. R.: *Regional Frequency Analysis. An approach based on L-moments*, Cambridge, Cambridge University Press, Cambridge, 1997.
- Hurst, H. E.: Long-Term storage capacity of reservoir, *Trans. Am. Soc. Civil Eng.*, 116, 770–799, 1951.
- Jonathan, P. and Ewans, K.: The effect of directionality on extreme wave design criteria, *Ocean Eng.*, 34, 1977–1994, 2007.
- Koscielny-Bunde, E., Kantelhardt, J. W., Braun, P., Bunde, A. and Havlin, S.: Long-term persistence and multifractality of river runoff records: Detrended fluctuation studies, *J. Hydrol. Hydrofractals '03*, 322, 120–137, 2006.
- Lang, M., Ouarda, T. B. M. J., and Bobée, B.: Towards operational guidelines for over-threshold modeling, *J. Hydrol.*, 225, 103–117, 1999.
- Leadbetter, M. R., Lindgren, G., and Rootzen, H.: *Extremes and Related Properties of Random Sequences and Processes*, New York, Springer-Verlag, 1983.
- Li, F., Bicknell, C., Lowry, R., and Li, Y.: A comparison of extreme wave sample analysis methods with 1994–2010 offshore Perth dataset., *Coast. Eng.*, 69, 1–11, 2012.
- Mackay, E. B. L., Challenor, P. G., and Bahaj, A. S.: A comparison of estimators for the generalised Pareto distribution, *Ocean Eng.*, 38, 1338–1346, 2001.
- MacKay, E. B. L., Challenor, P. G., and Bahaj, A. S.: On the use of discrete seasonal and directional models for the estimation of extreme wave conditions, *Ocean Eng.*, 37, 425–442, 2010.
- Madsen, H. and Rosbjerg, D.: The partial duration series method in regional index-flood modeling, *Water Resour. Res.*, 33, 737–746, 1997.
- Marani, M.: On the correlation structure of continuous and discrete point rainfall, *Water Resour. Res.*, 39, SWC21–SWC28, 2003.
- Mathiesen, M., Goda, Y., Hawkes, P. J., Mansard, E., Martin, M. J., Peltier, E., Thompson, E. F., and Van Vledder, G.: Recommended practice for extreme wave analysis, *J. Hydraul. Res.*, 32, 803–814, 1994.
- Mazas, F. and Hamm, L.: A multi-distribution approach to POT methods for determining extreme wave heights, *Coast. Eng.*, 58, 385–394, 2011.
- Neves, C. and Fraga Alves, M. I.: Reiss and Thomas' automatic selection of the number of extremes, *Computational Statistics & Data Analysis*, 47, 689–704, 2004.
- Ntegeka, V. and Willems, P.: Trends and multidecadal oscillation in rainfall extremes, based on a more than 100 years time series of 10 minutes rainfall intensities at Uccle, Belgium, *Water Resour. Res.*, 44, W07402, doi:10.1029/2007WR006471, 2008.
- Onoz, B. and Bayazit, M.: Effect of the occurrence process of the peaks over threshold on the flood estimates, *J. Hydrol.*, 244, 86–96, 2001.
- Ouarda, T. B. M. J., Rasmussen, P. F., Bobée, B., and Bernier, J.: Use of historical information in hydrology frequency analysis, *J. Water Sci.* 11, 41–49, 1998.
- Palutikof, J. P., Brabson, B. B., Lister, D. H., and Adcock, S. T.: A review of methods to calculate extreme wind speeds, *Meteorol. Appl.*, 6, 119–132, 1999.

- Pandey, M. D., Van Gelder, P. H. A. J. M., and Vrijling, J. K.: The estimation of extreme quantiles of wind velocity using L-moments in the peaks-over-threshold approach, *Struc. Saf.*, 23, 179–192, 2001.
- Parent, E. and Bernier, J.: Encoding prior experts judgments to improve risk analysis of extreme hydrological events via POT modeling, *J. Hydrol.*, 283, 1–18, 2003.
- Payraastre, O., Gaume, E., and Andrieu, H.: Use of historical data to assess the occurrence of floods in small watershed in the French Mediterranean area., *Adv. Geosci.*, 2, 313–320, 2005.
- Payraastre, O., Gaume, E., and Andrieu, H.: Usefulness of historical information for flood frequency analyses: Developments based on a case study, *Water Resour. Res.*, 47, W08511, doi:10.1029/2010WR009812, 2011.
- Pickands, J.: Statistical Inference Using Extreme Order Statistics, *The Annals of Statistics*, 3, 119–131, 1975.
- Ribatet, M., Sauquet, E., Grésillon, J.-M., and Ouarda, T. B. M. J.: A regional Bayesian POT model for flood frequency analysis, *Stoch. Environ. Res. Risk Assess.*, 21, 327–339, 2007.
- Rosbjerg, D.: Estimation in partial duration series with independent and dependent peak values, *J. Hydrol.*, 76, 183–195, 1985.
- Rosbjerg, D. and Madsen, H.: Advanced approaches in PDS/POT modelling of extreme hydrological events, *Hydrology – Science and Practise for the 21st Century*, B. H. Society, B. Webb., 217–220, 2004
- Rosbjerg, D., Madsen, H., and Rasmussen, P. F.: Prediction in partial duration series with generalized pareto-distributed exceedances, *Water Resour. Res.*, 28, 3001–3010, 1992.
- Ruggiero, P., Komar, P. D., and Allan, J. C.: Increasing wave heights and extreme value projections: The wave climate of the US Pacific Northwest, *Coast. Eng.*, 57, 539–552, 2010.
- Schertzer, D. and Lovejoy, S.: Universal multifractals do exist!: Comments, *J. of Appl. Meteorol.*, 36, 1296–1303, 1997.
- Silva, A. T., Portela, M. M., and Naghettini, M.: Nonstationarities in the occurrence rates of flood events in Portuguese watersheds, *Hydrol. Earth Syst. Sci.*, 8, 8609–8639, 2011.
- Simiu, E. and Heckert, A.: Extreme Wind Distribution Tails: A ‘Peaks Over Threshold’ Approach, National Science Foundation, 1995.
- Smith, O. P.: Duration of extreme wave conditions, *J. Waterw. Port C.-ASCE*, 114, 1–17, 1988.
- Smith, R. L.: Threshold methods for sampling extremes, *Statistical extremes and applications*, T. d. O. J. Dordrecht, teh Netherlands, 621–638, 1984
- Smith, R. L. and Weissman, I.: Estimating the Extremal Index, *J. R. Stat. Soc.*, 56, 515–528, 1994.
- Takvor, S. H. and Panagiota, A. M.: The Effect of Declustering in the r-Largest Maxima Model for the Estimation of Hs Design Values, *The Open Ocean Eng. J.*, 4, 34–43, 2001.
- Tawn, J. A. and Vassie, J. M.: Extreme sea level: the joint probability method revisited and revised, *Proc. Instn. Civ. Engrs.*, 429–442, 1989.
- Taylor, P. H., Barker, V. E., Bishop, D., and Eatock Taylor, R.: 100-year waves, teleconnections and wave climate variability, 11th International Workshop on Wave Hindcasting and Forecasting, Halifax, Canada, 2009.
- Thompson, P., Cai, Y., Reeve, D., and Stander, J.: Automated threshold selection methods for extreme wave analysis, *Coast. Eng.*, 56, 1013–1021, 2009.
- USWRC: Guidelines for determining flood flow frequency H. C. Bull.17, Washington DC, United States Water Resources Council, p. 73, 1976.
- Wahl, T., Mudersbach, C., and Jensen, J.: Assessing the hydrodynamic boundary conditions for risk analyses in coastal areas: a stochastic storm surge model, *Nat. Hazards Earth Syst. Sci.*, 11, 2925–2939, 2011.
- Walton, T. L.: Distribution for storm surge extremes, *Ocean Eng.*, 27, 1979–1293, 2000.
- Willems, P.: Hydrological applications of extreme value analysis, *Hydrology in a changing environment*, H. a. C. K. Wheeler. Chichester, John Wiley & Sons, 15–25, 1998
- Willems, P.: Compound intensity/duration/frequency-relationships of extreme precipitation for two seasons and two storm types, *J. Hydrol.*, 233, 189–205, 2000.
- Zawadzky, I.: Fractal structure and exponential decorrelation in rain, *J. Geophys. Res.*, 92, 1816–1818, 1987.