A horse racing between the block maxima method and the peak–over–threshold approach

Axel Bücher

Heinrich-Heine-Universität Düsseldorf, Mathematisches Institut, Universitätsstr. 1, 40225 Düsseldorf, Germany. e-mail: axel.buecher@hhu.de

and

Chen Zhou

Econometric Institute, Erasmus University Rotterdam, 3000 DR Rotterdam, The Netherlands. e-mail: zhou@ese.eur.nl; c.zhou@dnb.nl

Abstract: Classical extreme value statistics consists of two fundamental approaches: the block maxima (BM) method and the peak-over-threshold (POT) approach. It seems to be general consensus among researchers in the field that the POT method makes use of extreme observations more efficiently than the BM method. We shed light on this discussion from three different perspectives. First, based on recent theoretical results for the BM approach, we provide a theoretical comparison in i.i.d. scenarios. We argue that the data generating process may favour either one or the other approach. Second, if the underlying data possesses serial dependence, we argue that the choice of a method should be primarily guided by the ultimate statistical interest: for instance, POT is preferable for quantile estimation, while BM is preferable for return level estimation. Finally, we discuss the two approaches for multivariate observations and identify various open ends for future research.

Keywords and phrases: extreme value statistics, extreme value index, extremal index, stationary time series.

1. Introduction

Extreme-Value Statistics can be regarded as the art of extrapolation. Based on a finite sample from some distribution F, typical quantities of interest are quantiles whose levels are larger than the largest observation or probabilities of rare events which have not occurred yet in the observed sample. Estimating such objects typically relies on the following fundamental domain-of-attraction condition: there exists a constant $\gamma \in \mathbb{R}$ and sequences $a_r > 0$ and b_r, $r \in \mathbb{N}$, such that

$$\lim_{r\to\infty} F^r(a_r x + b_r) = \exp\left\{-(1+\gamma x)^{-1/\gamma}\right\} \text{ for all } 1+\gamma x > 0. \tag{1.1}$$

In that case, γ is called the *extreme value index*. The limit appears unnecessarily specific, but it is in fact the only non-degenerate limit of the expression on the left-hand side. An equivalent representation of the domain of attraction condition (1.1) is as follows: there exists a positive function $\sigma = \sigma(t)$ such that

$$\lim_{t\uparrow x^*} \frac{1-F(t+\sigma(t)x)}{1-F(t)} = (1+\gamma x)^{-1/\gamma} \text{ for all } 1+\gamma x > 0, \tag{1.2}$$

where x^* denotes the right end-point of the support of F, see Balkema and de Haan (1974). The two sequences in (1.1) are related to the function σ as follows: $a_r = \sigma(b_r)$ and $b_r = U(r)$ where $U(r) = F^{\leftarrow}(1-1/r) = (1/(1-F))^{\leftarrow}(r)$, with $\cdot^{\leftarrow}$ denoting the left–continuous inverse of some monotone function.

Consider for instance the consequences of the previous two displays for high quantiles of F. By by (1.1), for all p sufficiently small,

$$F^{\leftarrow}(1-p) \approx b_r + a_r \frac{\{-r\log(1-p)\}^{-\gamma}-1}{\gamma} \approx b_r + a_r \frac{(rp)^{-\gamma}-1}{\gamma}. \tag{1.3}$$

Hence, by the plug-in-principle, a suitable choice of r and suitable estimators of a_r, b_r and γ immediately suggest estimators for high quantiles.

Similarly, by (1.1), for all p sufficiently small,

$$F^{\leftarrow}(1-p) \approx t + \sigma(t) \frac{\{\frac{p}{1-F(t)}\}^{-\gamma}-1}{\gamma}. \tag{1.4}$$

Again, by the plug-in-principle, a suitable choice of t and suitable estimators of $\sigma(t)$, γ and $1-F(t)$ immediately leads to estimators for high quantiles. Here, t is typically chosen as a large order statistic $t = X_{n-k:n}$ and $1-F(t)$ is replaced by k/n.

In practice, estimators for the parameters in these two approaches typically follow their corresponding basic principles: the block maxima method motivated by (1.1) and the peak-over-threshold approach motivated by (1.2). Let $X_1, X_2, \dots, X_n$ be a sample of observations drawn from F, and for the moment assume that the observations are independent. Then (1.1) gives rise to the *block maxima method (BM)* (Gumbel, 1958): for some block size $r \in \{1,\dots,n\}$, divide the data into $k = \lfloor n/r \rfloor$ blocks of length r (and a possibly remaining block of smaller size which has to be discarded). By independence, each block has cdf F^r. By (1.1), for large block sizes r, the sample of block maxima can then be regarded as an approximate i.i.d. sample from the three-parametric generalized extreme-value (GEV) distribution $G_{\gamma,b,a}$ with location parameter $b = b_r$, scale parameter $a = a_r$ and shape parameter γ, defined by its cdf

$$G^{\mathrm{GEV}}_{\gamma,b,a}(x) := \exp\Big\{-\Big(1+\gamma\frac{x-b}{a}\Big)^{-1/\gamma}\Big\}\mathbf{1}\Big(1+\gamma\frac{x-b}{a}>0\Big).$$

The three parameters can be estimated by maximum-likelihood or moment-matching, among others. Irrespective of the particular estimation principle, any estimator defined in terms of the sample of block maxima will be referred to as an estimator based on the block maxima method.

Often, an available data-sample consists of block maxima only, for example, annual maxima of a river level. Then a practitioner may only rely on the block maxima method. If the underlying observations are available, then (1.2) gives rise to the competing *peak-over-threshold approach (POT)* (Pickands, 1975): for sufficiently large t in (1.2), we obtain that, for any $x > 0$,

$$\Pr(X > t+x \mid X > t) = \frac{\Pr(X>t+x)}{\Pr(X>t)} \approx \Big(1+\gamma\tfrac{x}{\sigma}\Big)^{-1/\gamma} =: 1 - G^{\mathrm{GP}}_{\gamma,\sigma}(x), \tag{1.5}$$

where the right-hand side defines the two-parametric generalized Pareto (GP) distribution with scale parameter $\sigma := \sigma(t)$ and shape parameter γ. In practice, t is typically chosen as the $(n-k)$-th order statistic $X_{n-k:n}$ for some intermediate value k (hence, $X_{n-k:n}$ is the $(1-1/r)$-sample quantile with $r = n/k$). Then, one may regard the sample $X_{n-k+1:n} - X_{n-k:n}, \dots X_{n:n} - X_{n-k:n}$ as observations from the two-parametric generalized Pareto-distribution. The parameters can hence be estimated by moment matching, and even by maximum-likelihood since the sample of order statistics can actually be regarded as independent (see, e.g., Lemma 3.4.1 in de Haan and Ferreira, 2006). In general, any estimator defined in terms of all observations exceeding some (random) threshold will be referred to as an estimator based on the POT approach. The vanilla estimator within this class is the Hill estimator (Hill, 1975) in the case $\gamma > 0$.

The goal of the present paper is an in-depth comparison of the two approaches, in particular in terms of recent solid theoretical advances on asymptotic theory for the BM method, but also with a view on time series data and multivariate observations. The discussion will mostly be of reviewing nature, but some new insights will be presented as well. The next paragraphs summarize our contribution in a chronological order.

1. **Efficiency comparison in i.i.d. scenarios.** It seems to be general consensus among researchers in extreme value statistics that the POT method produces more efficient estimators than the BM method. The main heuristic reason is that all large observations are used for the calculation of POT estimators, while BM estimators may miss some large observations falling into the same block. This heuristics was confirmed by simulation studies in Caires (2009), see also the additional references mentioned in Ferreira and de Haan (2015). Due to some recent advances on theoretic properties of BM estimators (Dombry, 2015; Ferreira and de Haan, 2015; Bücher and Segers, 2014, 2018b; Dombry and Ferreira, 2017), the two approaches may actually be compared on solid theoretical grounds. For a certain type of cdfs, such a discussion has been carried out in Ferreira and de Haan (2015) and Dombry and Ferreira (2017); their findings are summarized and extended in Section 2 of this paper. We show that, depending on the data generating process, the convergence rate of the two methods may be different, with no general winner being identifiable. In case the rates are the same, BM estimators typically have a smaller variance, but a larger bias than their POT-competitors.

2. **BM and POT applied to time series.** The above discussion motivating the BM and POT approach was based on an i.i.d. assumption on the underlying sample. This assumption is actually quite restrictive since it excludes many common environmental or financial applications, where the underlying sample is typically a (stationary) time series. In this setting, it seems to be general consensus that the block maxima method still 'works' because the block maxima are (1) still approximately GEV-distributed (Leadbetter, 1974) and (2) distant from each other and thus bear low serial dependence. Consequently, the sample of block maxima may still be regarded as an approximate i.i.d. sample from the three-parametric GEV-distribution. This heuristics is confirmed by recent theoretical results in Bücher and Segers (2018b, 2014).

Nevertheless, as discussed in Section 3 below, an obstacle occurs: the location and scale parameters attached to block maxima of a time series will typically be different from those of an i.i.d. series from the same stationary distribution F, whence estimators for quantities that depend on the stationary distribution only will possibly be inconsistent. The missing link is provided by the extremal index (Leadbetter, 1983), a parameter in $[0, 1]$ capturing the tendency of the extreme observations of a stationary time series to occur in clusters. The discussion will be worked out on the example of high quantile estimation: based on suitable estimators for the extremal index, see Section 3 below, (1.3) can in fact be modified to obtain consistent BM estimators of large quantiles.

On the other hand, estimators based on the POT method for characteristics of the stationary distribution remain consistent. This however comes at the cost of an increased variance of the estimators due to potential clustering of extremes, see Hsing, 1991; Drees, 2000; Rootzén, 2009, among many others. Should the ultimate interest be in return level or return periods estimation, the picture is reversed: the BM method is consistent without the need of estimating the extremal index, while POT estimators typically require estimates of the extremal index. More details are provided in Section 3.

3. **Extensions to multivariate observations and stochastic processes.** The previous discussion focussed on the univariate case. Section 4 briefly discusses multivariate extensions. On the theoretical side, while there are many results available for the POT approach, there is clearly a supply issue regarding the BM approach: almost all statistical theory is formulated under the

assumption that the block maxima genuinely follow a multivariate extreme value distribution, thereby ignoring a potential bias and rendering a fair theoretical comparison impossible for the moment (to the best of our knowledge, the only available results on the BM method are provided in Bücher and Segers, 2014). Instead, we provide a review on some of the existing theoretical results using these two approaches, and identify the open ends that may eventually lead to results allowing for an in-depth theoretical comparison in the future.

Not surprisingly, a fair comparison is even more difficult when considering extreme value analysis for stochastic processes. Most of the existing statistical methods are based on max-stable process models, i.e., on limit models arising for maxima taken over i.i.d. stochastic processes. The respective statistical theory is again mostly formulated under the assumption that the observations are genuine observations from the max-stable model, whence the statistical methods can (in most cases) be generically attributed to the BM approach. As for multivariate models, potential bias issues are mostly ignored. By contrast to multivariate models, however, very little is known for the POT approach to processes. A comparison is hence not feasible for the moment, and we limit ourselves to a brief review of existing results in Section 5.

Finally, we end the paper by a section summarizing possible open research questions, Section 6, and by a short conclusion, Section 7.

2. Efficiency Comparison for univariate i.i.d. observations

The efficiency of BM and POT estimators can be compared in terms of their asymptotic bias and variance. In this section, we particularly focus on the estimation of the extreme value index γ because for estimating other tail related characteristics such as high quantiles or tail probabilities, the asymptotic distributions of respective estimators are typically dominated by those derived from estimating the extreme value index.

In both the BM and POT approach, a key tuning parameter is the intermediate sequence $k = k(n)$, which corresponds to either the number of blocks in the BM approach, or the number of upper order statistics in the POT approach. For most data generating processes, consistency of respective estimators can be guaranteed if k is chosen in such a way that $k \to \infty$ and $k/n \to 0$ as $n \to \infty$. Here, the small fraction k/n reflects the fact that the inference is based on observations in the tail only. Typically, the variance of respective estimators is of order $1/k$, while the bias depends on how well the distribution of block maxima or threshold exceedances is approximated by the GEV or GP distribution, respectively. Choosing k in such a way that variance and squared bias are of the same order (see Section 2.1 below), one may derive an optimal rate of convergence for a given estimator. Depending on the model, the optimal choice of k may result in a faster rate for the BM method or the POT approach, as will be discussed next.

It is instructive to consider two extreme examples first (where the condition $k/n \to 0$ as $n \to \infty$ may in fact be discarded): if F is the standard Fréchet-distribution, then block maxima of size $r = 1$ are already GEV-distributed. In other words a sample of $k = n$ block maxima of size $r = 1$ can be used for estimation via the BM method. The rate of convergence is thus $1/\sqrt{n}$ and the POT method fails to achieve this rate. On the other hand, if F is the standard Pareto distribution, then all $k = n$ largest order statistics can be used for the estimation via the POT approach. The rate of convergence is $1/\sqrt{n}$ for the POT method, which is not achievable via the BM method.

Apart form these two (or similar) extreme cases, the optimal choice of k depends on second order conditions quantifying the speed of convergence in the domain of attraction condition.

These are often (though not always) formulated in terms of the two quantile functions

$$U(x) = \Big(\frac{1}{1-F}\Big)^{\leftarrow}(x) \quad \text{and} \quad V(x) = \Big(\frac{1}{-\log F}\Big)^{\leftarrow}(x)$$

for the POT- and the BM method, respectively. Note that the domain of attraction condition (1.1) is equivalent to the fact that there exists a positive function a_{POT} such that, for all $x > 0$,

$$\lim_{t\to\infty} \frac{U(tx) - U(t)}{a_{\text{POT}}(t)} = \int_1^x s^{\gamma-1}\, ds, \tag{2.1}$$

see Theorems 1.1.6 and 1.2.1 in de Haan and Ferreira (2006). The function a_{POT} is related to the sequence $(a_r)_r$ appearing in (1.1) via $a_{\text{POT}}(r) = a_{\lfloor r \rfloor}$.

In parallel, (1.1) is also equivalent to the fact there exists a positive function a_{BM} such that, for all $x > 0$,

$$\lim_{t\to\infty} \frac{V(tx) - V(t)}{a_{\text{BM}}(t)} = \int_1^x s^{\gamma-1}\, ds. \tag{2.2}$$

The bias of certain BM- and POT estimators is determined by the speed of convergence in the latter two limit relations, which can be captured by suitable second order conditions.

For $\gamma \in \mathbb{R}, \rho \le 0$ and $x > 0$, let

$$h_\gamma(x) = \int_1^x s^{\gamma-1}\, ds, \qquad H_{\gamma,\rho}(x) = \int_1^x s^{\gamma-1} \int_1^s u^{\rho-1}\, du\, ds.$$

Definition 2.1 (Second order conditions)**.** Let F be a cdf satisfying the domain-of-attraction condition (1.1) for some $\gamma \in \mathbb{R}$. Consider the following two assumptions.

$(\text{SO})_{\text{U}}$ Suppose that there exists $\rho_{\text{POT}} \le 0$, a positive function a_{POT} and a positive or negative function A_{POT} with $\lim_{t\to\infty} A_{\text{POT}}(t) = 0$, such that, for all $x > 0$,

$$\lim_{t\to\infty} \frac{1}{A_{\text{POT}}(t)} \Big(\frac{U(tx) - U(t)}{a_{\text{POT}}(t)} - h_\gamma(x) \Big) = H_{\gamma,\rho_{\text{POT}}}(x).$$

$(\text{SO})_{\text{V}}$ Suppose that there exists $\rho_{\text{BM}} \le 0$, a positive function a_{BM} and a positive or negative function A_{BM} with $\lim_{t\to\infty} A_{\text{BM}}(t) = 0$, such that, for all $x > 0$,

$$\lim_{t\to\infty} \frac{1}{A_{\text{BM}}(t)} \Big(\frac{V(tx) - V(t)}{a_{\text{BM}}(t)} - h_\gamma(x) \Big) = H_{\gamma,\rho_{\text{BM}}}(x).$$

The functions $|A_{\text{BM}}|$ and $|A_{\text{POT}}|$ are then necessarily regularly varying with index ρ_{BM} and ρ_{POT}, respectively. The limit function $H_{\gamma,\rho}$ might appear unnecessarily specific, but in fact it is not, see de Haan and Stadtmüller (1996) or Section B.3 in de Haan and Ferreira (2006). If the speed of convergence in (2.1) or (2.2) is faster than any power function, we set the respective second order parameter as minus infinity. For example, for $F = G^{\text{GP}}_{\gamma,\sigma}$ from the GP family, we have $\{U(tx) - U(t)\}/(\sigma t^\gamma) = h_\gamma(x)$, i.e. $\rho_{\text{POT}} = -\infty$ in this case. Likewise, any $F = G^{\text{GEV}}_{\gamma,\sigma,\mu}$ from the GEV distribution satisfies $\{V(tx) - V(t)\}/(\sigma t^\gamma) = h_\gamma(x)$, which prompts us to define $\rho_{\text{BM}} = -\infty$.

It is important to note that ρ_{BM} and ρ_{POT} can be vastly different. A general result can be found in Drees, de Haan and Li (2003), Corollary A.1: under an additional condition which only concerns the cases $\gamma = 1$, $\rho_{\text{BM}} = -1$ or $\rho_{\text{POT}} = -1$, the two coefficients are equal within the range $[-1, 0]$. Otherwise, if $(\text{SO})_{\text{V}}$ holds with $\rho_{\text{BM}} < -1$, then $(\text{SO})_{\text{U}}$ holds with $\rho_{\text{POT}} = -1$; if

Distribution	γ	ρ_{POT}	ρ_{BM}
$\mathrm{GP}(\gamma,\sigma)$	γ	$-\infty$	-1
Exponential(λ)	0	$-\infty$	-1
Uniform$(0,1)$	-1	$-\infty$	-1
Arcsin	-2	-2	-1
Burr(η,τ,λ)	$1/(\lambda\tau)$	$-1/\lambda$	$\max(-1/\lambda,-1)$
$t_\nu, \nu \neq 1$	$1/\nu$	$-2/\nu$	$\max(-2/\nu,-1)$
Cauchy$(=t_1)$	1	-2	-2
Weibull$(\lambda,\beta), \beta\neq 1$	0	0	0
$\Gamma(\alpha,\beta)$	0	0	0
Normal(μ,σ^2)	0	0	0
$F(x)=\exp(-(1+x^\alpha)^\beta)$	$1/(\alpha\beta)$	$\max(-1/\beta,-1)$	$-1/\beta$
Fréchet(α,σ)	$1/\alpha$	-1	$-\infty$
Reverse Weibull(β,μ,σ)	$-1/\beta$	-1	$-\infty$
GEV(γ,μ,σ)	γ	-1	$-\infty$

Table 1
Extreme value index and second order parameters for various models.

$(\mathrm{SO})_{\mathrm{U}}$ holds with $\rho_{\mathrm{POT}} < -1$, then $(\mathrm{SO})_{\mathrm{V}}$ holds with $\rho_{\mathrm{BM}} = -1$. Some values of the parameters for various types of distributions are collected in Table 1.

We remark that for the the t_1-distribution, we obtained $\rho_{\mathrm{BM}} = \rho_{POT} = -2$. This is a special example for which Corollary 4.1 in Drees, de Haan and Li (2003) is not applicable: $2tA(t)$ converges to $0 = 1-\gamma$. Notice that for the six models in the first category $\rho_{POT} < \rho_{BM}$ (if we consider $\lambda > 1$ in the Burr distribution and $\nu > 2$ in the t_ν distribution). For the four models in the second category $\rho_{POT} = \rho_{BM}$ while for the last four models, $\rho_{POT} > \rho_{BM}$ if we consider $\beta > 1$ in the model $F(x) = \exp(-(1+x^\alpha)^\beta)$.

Let us now consider asymptotic theory for the estimation of the extreme value index γ. Perhaps surprisingly, asymptotic theory for the BM method has hitherto mostly ignored the fact that block maxima are only approximately GEV distributed (see, e.g., Prescott and Walden, 1980; Hosking, Wallis and Wood, 1985; Bücher and Segers, 2017, among others). Only recent theoretical studies in Ferreira and de Haan (2015) and Dombry and Ferreira (2017) for the probability weighted moment (PWM) and the maximum likelihood (ML) estimator, respectively, take the approximation into account. Correspondingly, the asymptotic bias can be explicitly analyzed, relying on the second order condition $(\mathrm{SO})_{\mathrm{V}}$ above. On the other hand, solid theoretical studies regarding the POT method have a much longer history, see de Haan and Ferreira (2006) for a comprehensive overview. For the sake of theoretical comparability with the BM method, we will subsequently exemplarily deal with the ML estimator and the PWM estimator, for which Theorems 3.4.2 and 3.6.1 in de Haan and Ferreira (2006) provide the respective asymptotic theory under the assumption that $(\mathrm{SO})_{\mathrm{U}}$ is met (the results rely on Drees, 1998; Drees, Ferreira and de Haan, 2004).

Summarizing the above mentioned results, for both methods (BM and POT), the ML-estimators are consistent for $\gamma > -1$ and asymptotically normal for $\gamma > -1/2$, while PWM-estimators are consistent for $\gamma < 1$ and asymptotically normal for $\gamma < 1/2$. Asymptotic theory is formulated under the conditions that $k = k_n$ satisfies $k \to \infty$ and $k/n \to 0$ (POT method) or $r = r_n$ satisfies $r \to \infty$ and $k = r/n \to 0$ (BM method), as $n \to \infty$. Under the respective second order conditions $(\mathrm{SO})_{\mathrm{U}}$ and $(\mathrm{SO})_{\mathrm{V}}$ formulated above, the asymptotic results can be summarized as

$$\hat\gamma \overset{d}{\approx} \mathcal{N}\Big(\gamma + A_m(n/k)b, \frac{1}{k}\sigma^2\Big), \quad m \in \{\textsc{bm}, \textsc{pot}\},$$

where $\hat\gamma$ is one of the four estimators, and where the asymptotic bias b and the asymptotic variance σ^2 depend on the specific estimator, the second order index ρ_m and γ. In particular, the rate of convergence of the bias $A_m(n/k)$ crucially depends on the second order index ρ_m.

In the next two subsections, we first discuss the best achievable rate of convergence and then the asymptotic mean squared error in case the rates are the same. Finally, in the last subsection, we discuss the choice of k, i.e., the number of large order statistics in the POT approach or the number of blocks in the BM approach.

2.1. Rate of convergence

As is commonly done, we consider the rate of convergence of the root mean squared error. It is instructive to first elaborate on the case $A_m(t) \asymp t^{\rho_m}$ with $\rho_m \in (-\infty, 0)$. The best attainable rate of convergence is achieved when squared bias and variance are of the same order, that is, when

$$A_m^2\left(\frac{n}{k}\right) \asymp \left(\frac{n}{k}\right)^{2\rho_m} \asymp \frac{1}{k}.$$

Solving for k yields $k \asymp n^{-2\rho_m/(1-2\rho_m)}$, which implies

$$\text{Rate of Convergence of } \hat{\gamma} = n^{\rho_m/(1-2\rho_m)}$$

irrespective of $m \in \{\text{ML}, \text{PWM}\}$. For the POT approach, this result is known to hold for many other estimators of γ; see de Haan and Ferreira (2006) (though not for every estimator, see Table 3.1 in that reference). In fact, it can be shown that this is the optimal rate under some specific assumptions on the data generating process, see Hall and Welsh (1984). We conjecture that the same result holds true for many other estimators relying on the BM method, though the literature does not provide sufficient theoretical results yet except for the ML and PWM estimators.

Since ρ_{BM} and ρ_{POT} might not be the same, the best attainable rate of convergence may be different for the BM and POT approach. Table 2 provides a summary of which method results in a better rate. The case where the rates are the same is discussed in more detail in Section 2.2 below.

2nd Order Parameters	Rate POT	Rate BM	Better rate
$\rho = \rho_{\text{BM}} = \rho_{\text{POT}} \in [-1, 0)$	$n^{\rho/(1-2\rho)}$	$n^{\rho/(1-2\rho)}$	-
$\rho_{\text{BM}} = -1, \rho_{\text{POT}} < -1$	$n^{\rho_{\text{POT}}/(1-2\rho_{\text{POT}})}$	$n^{-1/3}$	POT
$\rho_{\text{POT}} = -1, \rho_{\text{BM}} < -1$	$n^{-1/3}$	$n^{\rho_{\text{BM}}/(1-2\rho_{\text{BM}})}$	BM

Table 2
Best attainable convergence rates for the BM and POT approach in case $A_m(t) \asymp t^{\rho_m}$ with $\rho_m < 0$ and for typical relationships between ρ_{BM} and ρ_{POT} (see Drees, de Haan and Li, 2003).

Let us finally mention that the specific assumption on the function A_m made above (i.e., $A_m(t) \asymp t^{\rho_m}$ with $\rho_m \in (-\infty, 0)$) is not essential, see the argumentation on pages 79–80 in de Haan and Ferreira (2006). Moreover, for $\rho_m = -\infty$, the convergence rate is 'faster than $n^{-1/2+\varepsilon}$ for any $\varepsilon > 0$', and, depending on the underlying distribution, in fact could even achieve $n^{-1/2}$ (see also Remark 3.2.6 in de Haan and Ferreira, 2006).

2.2. Asymptotic mean squared error

As discussed in the previous subsection, if $\rho_{\text{POT}} \neq \rho_{\text{BM}}$, the approach corresponding to a lower ρ generically yields estimators for γ with a faster attainable rate of convergence than the other approach. In this subsection, we consider the case $\rho_{\text{POT}} = \rho_{\text{BM}}$. Then both approaches, at their best attainable rate of convergence, will yield estimators of γ with the same speed of convergence.

Hence, the efficiency comparison should be made at the level of asymptotic mean squared error (AMSE) or, more precisely, its two subcomponents: asymptotic bias and asymptotic variance. Notice that the asymptotic bias and variance depends on the specific estimator used, whence the comparison can only be performed based on some preselected estimators.

A detailed analysis of the PWM and the ML estimators under the BM and POT approach has been carried out in Ferreira and de Haan (2015) and Dombry and Ferreira (2017), for the case $\rho_{\mathrm{BM}} = \rho_{\mathrm{POT}} \in [-1, 0]$ and $\gamma \in (-0.5, 0.5)$. The results are as follows: when using the same value for k, being either the number of large order statistics in the POT approach or the number of blocks in the BM approach, the BM version of either ML or PWM leads to a lower asymptotic variance compared to the corresponding POT version, for all $\gamma \in (-0.5, 0.5)$. On the other hand, the (absolute) asymptotic bias is smaller for the POT versions of the two estimators, for all $(\gamma, \rho) \in (-0.5, 0.5) \times [-1, 0]$.

When comparing the optimal AMSE (where optimal refers to the fact that the parameter k is chosen in such a way that the AMSE for the specific estimator is minimized), it turns out that, for the ML estimator, the POT approach yields a smaller optimal AMSE. For the PWM estimator, the BM method is preferable for most combinations of (γ, ρ). When comparing all four estimators, the combination ML-POT has the overall smallest optimal AMSE.

2.3. Threshold and block length choice

Both the POT and the BM approach require a practical selection for the intermediate sequence $k = k_n$ in a sample of size n. In the POT approach, the choice of k problem can be interpreted as the choice of the threshold above which the POT approximation in (1.5) is regarded as sufficiently accurate. Similarly, in the BM approach, k is related to $r = n/k$, which is the size of the block of which the GEV approximation to the block maximum is regarded as sufficiently accurate.

The theoretical conditions that $k \to \infty$ and $k/n \to 0$, as $n \to \infty$ are useless in guiding the practical choice. Practically, often a plot between the estimates based on various k against the values of k is made for resolving this problem, the so-called "Hill plot" (Drees, de Haan and Resnick, 2000), despite the fact that it can be also be applied to other POT or even BM estimators than just the Hill estimator. The ultimate choice is then made by taking a k from the first stable region in the "Hill plot". Nevertheless, the estimators are often rather sensitive to the choice of k.

For the POT approach, there exist a few attempts on resolving the choice of k issue in a formal manner. For example, one solution is to find the optimal k that minimizes the asymptotic MSE; see, e.g., Danielsson et al. (2001), Drees and Kaufmann (1998) and Guillou and Hall (2001). As an indirect solution to the problem, one may also rely on bias corrections, which typically allows for a much larger choice of k, see, e.g., Gomes, De Haan and Rodrigues (2008). After the bias correction, the "Hill plot" usually shows a stable behavior and the estimates are less sensitive to the choice of k. For an extensive review on bias corrections, see Beirlant, Caeiro and Gomes (2012).

Compared to the extensive studies on the threshold choice and on bias corrections for the POT approach, there is, to the best of our knowledge, no existing literature addressing these issues for the BM approach. This may partly be explained by the fact that block sizes are often given by the problem at hand, for instance, block sizes corresponding to year. Nevertheless, based on the recent solid theoretical advances on the BM method, the foundations are laid to explore these issues in a rigorous manner in the future.

3. BM and POT for Univariate Stationary Time Series

In many practical applications, the discussion from the previous section is not quite helpful: the underlying data sample is not i.i.d., but in fact a stretch of a possibly non-stationary time series. Often, by either restricting attention to a proper time horizon or by some suitable transformation, the time series can at least be assumed to be stationary.[1] Throughout this section, we make the following generic assumption: $(X_t)_{t\in\mathbb{Z}}$ is a strictly stationary univariate time series, and the stationary cdf F satisfies the domain-of-attraction condition (1.1). It is important to note that the parameters γ, a_r and b_r only depend on the stationary cdf F, and that for instance (1.3) expressing high quantiles of F through these parameters continues to hold for time series. Let us begin by passing over the arguments from Section 1 that eventually led to the BM- and POT method.

3.1. The POT approach for time series

Recall that the POT approach is based on the sample of large order statistics denoted by $\mathcal{X}_{\text{POT}} = \{X_{n-k:n}, \ldots, X_{n:n}\}$. The main motivation that lead us to consider this sample was the marginal limit relation (1.2). Bearing in mind that, under mild extra conditions on the serial dependence (ergodicity, mixing conditions, ...), empirical moments are consistent for their theoretical counterparts, it is thus still reasonable to estimate the respective parameters by any form of moment matching, e.g., by PWM. The asymptotic variance of such estimators will however be different from the i.i.d. case in general (a consequence of central limit theorems for time series under mixing conditions).

Consider the ML-method: unlike for i.i.d. data, the sample $\mathcal{X}_{\text{POT}}$ cannot be regarded as independent anymore, whence it is in general impossible to derive the (approximate) generalized Pareto likelihood of $\mathcal{X}_{\text{POT}}$. As a circumvent, one may 'do as if' the likelihood arising in the i.i.d. case is also the likelihood for the time series case (quasi-maximum likelihood), and use essentially the same ML-estimators as for the i.i.d. case. Then, since the latter estimator is in fact also depending on empirical moments only, we still obtain proper asymptotic properties such as consistency and asymptotic normality.

Respective theory can be found in Hsing (1991); Resnick and Stărică (1998) for the Hill estimator and in Drees (2000) for a large class of estimators, including PWM and ML. Most of the estimators have the same bias as in the i.i.d. case, whereas their asymptotic variances depend on the serial dependence structure and are usually higher than those obtained in the i.i.d. case. Since the asymptotic bias shares the same explicit form, bias correction can also be performed in the same way as in the i.i.d. case; see, e.g., De Haan, Mercadier and Zhou (2016).

3.2. The BM approach for time series

Recall that the BM approach is based on the sample of block maxima $\mathcal{X}_{\text{BM}} = \{M_{1,r}, \ldots, M_{k:r}\}$, where $M_{j,r}$ denotes the maximum within the jth disjoint block of observations of size r. The main motivation in Section 1 that lead us to consider this sample as approximately GEV-distributed was the relation

$$\Pr(M_{1,r} \le a_r x + b_r) = F^r(a_r x + b_r) \approx G^{\text{GEV}}_{\gamma,0,1}(x),$$

[1] For example, for financial applications, the stationarity assumption can often be approximately guaranteed by restricting attention to a time horizon during which few macro economic conditions had changed. Similarly, for environmental applications, this can be achieved by restricting attention to observations falling into, say, the summer months.

for large r. The first equality is not true for time series, whence more sophisticated arguments must be found for the BM method to work for time series. In fact, it can be shown that if F satisfies (1.1), if $\Pr(M_{1,r} \le a_r x + b_r)$ is convergent for some x and if mild mixing conditions on the serial dependence (known as $D(u_n)$-conditions) are met, then there exists a constant $\theta \in [0, 1]$ such that

$$\lim_{r\to\infty} \Pr(M_{1:r} \le a_r x + b_r) = \big(G^{\text{GEV}}_{\gamma,0,1}(x)\big)^{\theta}$$

for all $x \in \mathbb{R}$ (Leadbetter, 1983). The constant θ is called the *extremal index* and can be interpreted as capturing the tendency of the time series that extremal observations occur in clusters. If $\theta > 0$, then letting

$$\tilde{a}_r = a_r \theta^{\gamma}, \qquad \tilde{b}_r = b_r - a_r \frac{1-\theta^{\gamma}}{\gamma} \tag{3.1}$$

we immediately obtain that

$$\lim_{r\to\infty} \Pr(M_{1:r} \le \tilde{a}_r x + \tilde{b}_r) = G^{\text{GEV}}_{\gamma,0,1}(x) \tag{3.2}$$

for all $x \in \mathbb{R}$. Hence, the sample $\mathcal{X}_{\text{BM}}$ is approximately GEV-distributed with parameter $(\tilde{a}_r, \tilde{b}_r, \gamma)$, which can then be estimated by any method of choice. It is important to note that, unless $\theta = 1$ or $\gamma = 0$, $\tilde{a}_r$ and $\tilde{b}_r$ are different from a_r and b_r. Consequently, additional steps must be taken for estimating quantiles of F via (1.3), see also Section 3.3.2 below. Via (3.1), it is possible to transform between (a_r, b_r) and $(\tilde{a}_r, \tilde{b}_r)$ if the extremal index θ is known or estimated. Regarding the estimation of the extremal index, a large variety of estimators has been proposed, which may itself be grouped into four categories: 1) BM-like estimators based on "blocking" techniques (Northrop, 2015; Berghaus and Bücher, 2017), 2) POT-like estimators that rely on threshold exceedances (Ferro and Segers, 2003; Süveges, 2007), 3) estimators that use both principles simultaneously (Hsing, 1993; Robert, 2009; Robert, Segers and Ferro, 2009) and 4) estimators which, next to choosing a threshold sequence, require the choice of a *run-length* parameter (Smith and Weissman, 1994; Weissman and Novak, 1998).

Since the distance between the time points at which the maxima within two successive blocks are attained is likely to be quite large, the sample $\mathcal{X}_{\text{BM}}$ can be regarded as approximately independent. As a matter of fact, the literature on statistical theory for the BM method is mostly based on the assumption that $\mathcal{X}_{\text{BM}}$ is a genuine i.i.d. sample from the GEV-family (see, e.g., Prescott and Walden, 1980; Hosking, Wallis and Wood, 1985; Bücher and Segers, 2017, among others). Two approximation errors are thereby completely ignored: the cdf is only approximately GEV, and the sample is only approximately independent. Solid theoretical results taking these errors into account are rare: Bücher and Segers (2018b) treat the ML-estimator in the heavy-tailed case ($\gamma > 0$). The main conclusions are: the sample can safely be regarded as independent, but a bias term may appear which, similar as in Section 2, depends on the speed of convergence in (3.2). Bücher and Segers (2018a) improve upon that estimator by using sliding blocks instead of disjoint blocks. The asymptotic variance of the estimator decreases, while the bias stays the same. Moreover, the resulting 'Hill-Plots' are much smoother, guiding a simpler choice for the block length parameter.

3.3. Comparison between the two methods

Let us summarize the main conceptual differences between the BM and the POT method for time series. First of all, BM and POT estimate 'the same' extreme value index γ, but possibly different scaling sequence $\tilde{a}_r, \tilde{b}_r$ and a_r, b_r. Second, the sample $\mathcal{X}_{\text{BM}}$ can be regarded as asymptotically independent (asymptotic variances of estimators are the same as if the sample was i.i.d.), while $\mathcal{X}_{\mathrm{POT}}$ is serially dependent, possibly increasing asymptotic variances of estimators compared to the i.i.d. case.

Due to the lack of a general theoretical result on the BM method, a theoretical comparison on which method is more efficient along the lines of Section 2 seems out of reach for the moment. In particular, a relationship between the respective second order conditions controlling the bias is yet to be found. However, some insight into the merits and pitfalls of two approaches can be gained by considering the problem of estimating high quantiles and return levels.

3.3.1. Estimating high quantiles

Recall that high quantiles of the stationary distribution can be expressed in terms of a_r, b_r and γ, see (1.3). As a consequence, based on the plug-in principle, the POT method immediately yields estimators for high quantiles. On the other hand, the BM method cannot be used straightforwardly, as it commonly only provides estimators of $\tilde{a}_r, \tilde{b}_r$ and γ. Via (3.1), the latter estimators may be transferred into estimators of a_r, b_r and γ using an additional estimator of the extremal index θ. It is important to note that the latter estimators typically depend on the choice of one or two additional parameters, and that they are often quite variable. By contrast, the POT approach therefore seems more suitable when estimating high quantiles or, more generally, parameters that only depend on the stationary distribution (such as probabilities of rare events). Recall though that estimators based on the POT approach usually suffer from a higher asymptotic variance due to the serial dependence.

3.3.2. Estimating return levels

Let $F_r(x) = \Pr(M_{1:r} \le x)$. For $T \ge 1$, the T-return level of the sequence of block maxima is defined as the $1 - 1/T$ quantile of F_r, that is,

$$\mathrm{RL}(T, r) = F_r^{\leftarrow}(1 - 1/T) = \inf\{x \in \mathbb{R} : F_r(x) \ge 1 - 1/T\}.$$

Since block maxima are asymptotically independent, it will take on average T blocks of size r until the first such block whose maximum exceeds $\mathrm{RL}(T, r)$. Now, since F_r is approximately equal to the GEV-cdf with parameters $\gamma, \tilde{b}_r, \tilde{a}_r$ for large r by (3.2), we obtain that

$$\mathrm{RL}(T, r) \approx \tilde{b}_r + \tilde{a}_r \frac{\{-r\log(1 - 1/T)\}^{-\gamma} - 1}{\gamma} \approx \tilde{b}_r + \tilde{a}_r \frac{(r/T)^{-\gamma} - 1}{\gamma}.$$

In comparison to the estimation of high-quantiles, see (1.3), we have now expressed the object of interest in terms of the sequences $\tilde{a}_r$ and $\tilde{b}_r$ and the extreme-value index γ. Following the discussion in the previous section, it is now the BM method which yields simpler estimators that do not require additional estimation of the extremal index. By contrast, the POT approach only results in estimators of (a_r, b_r) and γ, and therefore requires a transformation to $(\tilde{a}_r, \tilde{b}_r)$ via (3.1) based on an estimate of the extremal index θ.

4. BM and POT for Multivariate Observations

Due to the lack of asymptotic results on the multivariate BM method which take the approximation error into account, a deep comparison between the BM and POT approach is not feasible yet. Within this section we try to identify the open ends that may eventually lead to such results in the future.

Let F be a d-dimensional cdf. The basic assumption of multivariate extreme-value theory, generalizing (1.1), is as follows: suppose that there exists a non-degenerate cdf G and sequences $(a_{r,j})_{r\in\mathbb{N}}, (b_{r,j})_{r\in\mathbb{N}}, j = 1, \dots d$, with $a_{r,j} > 0$ such that

$$\lim_{r\to\infty} \Pr\Big(\frac{\max_{i=1}^r X_{i,1} - b_{r,1}}{a_{r,1}} \le x_1, \dots, \frac{\max_{i=1}^r X_{i,d} - b_{r,d}}{a_{r,d}} \le x_d\Big) = G(x_1, \dots, x_d) \tag{4.1}$$

for any $x_1, \dots, x_d \in \mathbb{R}$, where $\boldsymbol{X}_i = (X_{i,1}, \dots, X_{i,d})', i \in \mathbb{N}$, is an i.i.d. sequence from F, and where the marginal distributions G_j of G, $j = 1, \dots, d$, are GEV-distributions with location parameter 0, scale parameter 1 and shape parameter $\gamma_j \in \mathbb{R}$ (location 0 and scale 1 can always be reached by adapting the sequences $a_{r,j}$ are $b_{r,j}$ if necessary). The dependence between the coordinates of G can be described in various equivalent ways (see, e.g., Resnick, 1987; Beirlant et al., 2004; de Haan and Ferreira, 2006): by the stable tail dependence function L (Huang, 1992), by the exponent measure μ (Balkema and Resnick, 1977), by the Pickands dependence function A (Pickands, 1981), by the tail copula Λ (Schmidt and Stadtmüller, 2006), by the spectral measure Φ (de Haan and Resnick, 1977), by the madogram ν (Naveau et al., 2009), or by other less popular objects. All these objects are in one-to-one correspondence, and for each of them a large variety of estimators has been proposed, both in a nonparametric way and under the assumption that the objects are parametrized by an Euclidean parameter.

In this paper, we will mainly focus on nonparametric estimation. As in the univariate case, the estimators may again be grouped into BM and POT based estimators, see Sections 4.1 and 4.2 below. Often, estimation of the marginal parameters and of the dependence structure is treated successively. It is important to note that standard errors for estimators of the dependence structure may then be influenced by standard errors for the marginal estimation, a point which is often ignored in the literature on statistics for multivariate extremes. In fact, a phenomenon well-known in statistics for copulas (Genest and Segers, 2010) may show up: possibly completely ignoring additional information about the marginal cdfs, estimators for the dependence structure may have a lower asymptotic variance if marginal cdfs are estimated nonparametrically; see Bücher (2014) for a discussion of the empirical stable tail dependence function from Section 4.1 below, and Genest and Segers (2009) for estimation of Pickands dependence function based on i.i.d. data from a bivariate extreme value distribution, Section 4.2 below.

4.1. The POT method in the multivariate case

Suppose $\boldsymbol{X}_1, \dots, \boldsymbol{X}_n$, with $\boldsymbol{X}_i = (X_{i,1}, \dots, X_{i,d})'$, is an i.i.d. sample from F. Recall that the univariate POT method was based on the observations $\mathcal{X}_{\text{POT}} = \{X_{n-k:n}, \dots, X_{n:n}\}$, which may be rewritten as $\mathcal{X}_{\text{POT}} = \{X_i : \text{rank}(X_i \text{ among } X_1, \dots, X_n) \ge n-k)$. Thus, a possible generalization to multivariate observations consists of defining

$$\mathcal{X}_{\text{POT}} = \{\boldsymbol{X}_i \mid \text{rank}(X_{i,j} \text{ among } X_{1,j}, \dots, X_{n,j}) \ge n-k \text{ for some } j = 1, \dots, d\},$$

that is, $\mathcal{X}_{\text{POT}}$ comprises all observations for which at least one coordinate is large. Any estimator defined in terms of these observations may be called an estimator based on the multivariate POT method.

As an example, consider the estimation of the so-called stable tail dependence function L, which is defined as

$$L(\boldsymbol{x}) = \lim_{t\downarrow 0} t^{-1} \Pr(F_1(X_1) > 1 - tx_1 \text{ or } \dots \text{ or } F_d(X_d) > 1 - tx_d), \tag{4.2}$$

where $\boldsymbol{x} = (x_1, \ldots, x_d)' \in [0,1]^d$; a limit that necessarily exists under (4.1), but may also exist for marginals F_j not in any domain-of-attraction. The function L can be estimated by its empirical counterpart, defined as

$$\hat{L}(x_1, \ldots, x_d) = \frac{1}{k} \sum_{i=1}^{n} \mathbf{1}(\hat{F}_{n,1}(X_{i,1}) > 1 - \tfrac{k}{n} x_1 \text{ or } \ldots \text{ or } \hat{F}_{n,d}(X_{i,d}) > 1 - \tfrac{k}{n} x_d),$$

where $\hat{F}_{n,j}$ denotes the empirical cdf based on the observations $X_{1,j}, \ldots, X_{n,j}$; see, e.g., Huang (1992). Since $\boldsymbol{x} \in [0,1]^d$, the estimator in fact only depends on the sample $\mathcal{X}_{\mathrm{POT}}$.

Suppose the following natural second order condition quantifying the speed of convergence in (4.2) is met: there exists a positive or negative function A and a real-valued function $g \not\equiv 0$ such that

$$\lim_{t \to \infty} \frac{t \Pr(F_1(X_1) > 1 - \frac{x_1}{t} \text{ or } \ldots \text{ or } F_d(X_d) > 1 - \frac{x_d}{t}) - L(x_1, \ldots, x_d)}{A(t)} = g(\boldsymbol{x}) \tag{4.3}$$

uniformly in $\boldsymbol{x} \in [0,1]^d$. Then, under additional smoothness conditions on L, it can be shown that $\hat{L}$ is consistent and asymptotically Gaussian in terms of functional weak convergence, the variance being of order $1/k$ and the bias being of order $A(n/k)$, provided that $k = k_n \to \infty$ and $k/n \to 0$ as $n \to \infty$; see, e.g., Huang (1992); Einmahl, Krajina and Segers (2012), among others. Following the discussion in Section 2, if we additionally assume that $A(t) \asymp t^\rho$ for some $\rho \in (-\infty, 0)$, the best attainable convergence rate, achieved when squared bias and variance are balanced, is

$$\text{Rate of Convergence of } \hat{L}(\boldsymbol{x}) = n^{\rho/(1-2\rho)}.$$

This convergence rate is in fact optimal under additional conditions on the data-generating process, see Drees and Huang (1998). Also note that $\hat{L}$ suffers from an asymptotic bias as in the univariate case, and that corresponding bias corrections for the bivariate case have been proposed in Fougères et al. (2015).

As in the univariate case, the literature on further theoretical foundations for the multivariate POT method is vast, see, e.g., Einmahl, de Haan and Piterbarg (2001); Einmahl and Segers (2009) for nonparametric estimation of the spectral measure, Drees and de Haan (2015) for estimation of failure probabilities, or de Haan, Neves and Peng (2008); Einmahl, Krajina and Segers (2012) for parametric estimators, among many others.

4.2. The BM method in the multivariate case

Again suppose $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n$ is an i.i.d. sample from F. Let r denote a block size, and $k = \lfloor n/r \rfloor$ the number of blocks. For $\ell = 1, \ldots, k$, let $\boldsymbol{M}_{\ell,r} = (M_{\ell,1,r}, \ldots, M_{\ell,1,r})'$ denote the vector of componentwise block-maxima in the ℓth block of observations of size r (it is worthwhile to note that $\boldsymbol{M}_{\ell,r}$ may be different from any $\boldsymbol{X}_i$). Any estimator defined in terms of the sample $\mathcal{X}_{\mathrm{BM}} = (\boldsymbol{M}_{1,r}, \ldots, \boldsymbol{M}_{k,r})$ is called an estimator based on the BM approach.

Just as for the univariate BM method, asymptotic theory is usually formulated under the assumption that $\boldsymbol{M}_1, \ldots, \boldsymbol{M}_k$ is a genuine i.i.d. sample from the limiting distribution G; a potential bias is completely ignored. Moreover, estimation of the marginal parameters is often disentangled from estimation of the dependence structure, with theory for the latter either developed under the assumption that marginals are completely known (which usually leads to wrong asymptotic variances), or under the assumption that marginals are estimated nonparametrically. See, for instance, Pickands (1981); Capéraà, Fougères and Genest (1997); Zhang, Wells and Peng (2008); Genest and Segers (2009); Gudendorf and Segers (2012) for nonparametric estimators and Genest, Ghoudi and Rivest (1995); Dombry, Engelke and Oesting (2016) for parametric ones, among many others.

To the best of our knowledge, the only reference that takes the approximation error induced by the assumption of observing data from a genuine extreme-value model into account is Bücher and Segers (2014), where the estimation of the Pickands dependence function A based on the BM-method is considered. Not only the bias is treated carefully there, but also the underlying observations $\boldsymbol{X}_1, \dots, \boldsymbol{X}_n$ may possess serial dependence in form of a stationary time series. Just like in the univariate case described above, the best attainable convergence rate of the estimator again depends on a second order condition.

4.3. Comparison between the two methods

Due to the lack of honest theoretical results on the BM method, not much can be said yet about which method is better in terms of, say, the rate of convergence. The missing tool is a multivariate version of Corollary A.1 in Drees, de Haan and Li (2003), allowing one to move from a BM second order condition (such as the one imposed in Bücher and Segers, 2014) to a POT second order condition as in (4.3), and vice versa. It then seems likely that similar phenomena as in the univariate case in Section 2 may show up.

4.4. Multivariate time series

Moving from i.i.d. multivariate observations to multivariate strictly stationary time series induces similar phenomena as in the univariate case, whence we keep the discussion quite short. Under suitable conditions on the serial dependence, estimators based on the POT approach are still consistent and asymptotically normal, though with a possibly different asymptotic variance (this can for instance be deduced from Drees and Rootzén, 2010). Regarding the BM method, the same heuristics as in the univariate case apply: block maxima may safely be assumed as independent and as following a multivariate extreme value distribution (Bücher and Segers, 2014). The estimators based on the BM method are then also consistent and asymptotically normal with a potential bias. Similar to the discussion on the location and scale parameters in the univariate case, the objects that are estimated by POT and BM may be different but are linked by the multivariate extremal index (Nandagopalan, 1994, see also Section 10.5 in Beirlant et al., 2004). Hence, following the discussion in Section 3.3, it seems preferable to estimate quantities that only depend on the tail of the stationary distribution by the POT approach, while tail quantities similar to the univariate return levels (that also depend on the serial dependence) are preferably estimated by the BM approach. As in the univariate case, a detailed theoretical comparison does not seem to be feasible.

5. BM and POT for stochastic processes

The BM approach for stochastic processes is based on modeling by max-stable processes, i.e., on limit models arising for block maxima taken over i.i.d. stochastic processes. Recent research has focussed on the structure and characteristics of max-stable processes, see, e.g., De Haan (1984), Giné, Hahn and Vatan (1990) and Kabluchko, Schlather and De Haan (2009); on simulating from max-stable processes, see, e.g., Dombry, Éyi-Minko and Ribatet (2013), Dieker and Mikosch (2015), Dombry, Engelke and Oesting (2016) and Oesting, Schlather and Zhou (2018); and on statistical inference based on max-stable processes, see, e.g., Coles and Tawn (1996), Buishand, De Haan and Zhou (2008), Padoan, Ribatet and Sisson (2010) and Huser and Davison (2014).

As mentioned in the introduction, there is a clear supply issue regarding the POT approach to stochastic process models. Early studies such as Einmahl and Lin (2006) consider the estimation of marginal parameters only, or consider nonparametric estimation of the dependence structure (de Haan and Lin, 2003), however with only weak consistency established. Recent development on *Generalized Pareto Processes* allow for considering parametric estimation for the dependence structure, see, e.g. Ferreira and De Haan (2014), Thibaud and Opitz (2015) and Huser and Wadsworth (2017). Given the imbalanced nature, we skip a deeper review on the BM and POT approaches for extremes regarding stochastic processes.

6. Open problems

Throughout this paper, we have already identified a number of open research problems, mostly related to an honest verification of the BM approach. Within the following list, we recapitulate those issues and add several further possible research questions:

- Asymptotic theory on further estimators based on the block maxima method, if possible allowing for a comparison between the imposed second order condition and those from the POT approach.
- In case the BM method yields to faster attainable rates of convergence than the POT approach (Section 2.1): are the obtained rates optimal?
- Derive a test for which approach is preferably for a given data set ($H_0 : \rho_{\mathrm{BM}} \le \rho_{\mathrm{POT}}$, or similar).
- Block length choice and bias reduction for BM.
- More results on the sliding block maxima method (non-heavy tailed case, multivariate case).
- A comparison of BM and POT second order conditions in the multivariate case.
- A comparison of return level/quantile estimation based on BM and POT, possibly incorporating an estimator for the extremal index.
- Extension to stochastic processes (max-stable processes and generalized Pareto processes): theoretical results on statistical methodology are still rare, and a comparison between BM and POT is not feasible yet.

7. Conclusion

There is no winner.

References

Balkema, A. A. and de Haan, L. (1974). Residual life time at great age. *The Annals of Probability* **2** 792–804.

Balkema, A. A. and Resnick, S. I. (1977). Max-infinite divisibility. *J. Appl. Probability* **14** 309–319. MR0438425

Beirlant, J., Caeiro, F. and Gomes, M. I. (2012). An overview and open research topics in statistics of univariate extremes. *Revstat* **10** 1–31.

Beirlant, J., Goegebeur, Y., Segers, J. and Teugels, J. (2004). *Statistics of extremes: Theory and Applications*. *Wiley Series in Probability and Statistics*. John Wiley & Sons Ltd., Chichester.

Berghaus, B. and Bücher, A. (2017). Weak convergence of a pseudo maximum likelihood estimator for the extremal index. ArXiv e-prints, arXiv:1608.01903.

Bücher, A. (2014). A note on nonparametric estimation of bivariate tail dependence. *Stat. Risk Model.* **31** 151–162. . MR3213603

Bücher, A. and Segers, J. (2014). Extreme value copula estimation based on block maxima of a multivariate stationary time series. *Extremes* **17** 495–528.

Bücher, A. and Segers, J. (2017). On the maximum likelihood estimator for the generalized extreme-value distribution. *Extremes* **20** 839–872. . MR3737387

Bücher, A. and Segers, J. (2018a). Inference for heavy tailed stationary time series based on sliding blocks. *Electron. J. Statist.* **12** 1098–1125.

Bücher, A. and Segers, J. (2018b). Maximum likelihood estimation for the Fréchet distribution based on block maxima extracted from a time series. *Bernoulli* **24** 1427–1462. . MR3706798

Buishand, T., De Haan, L. and Zhou, C. (2008). On spatial extremes: with application to a rainfall problem. *The Annals of Applied Statistics* **2** 624–642.

Caires, S. (2009). A comparative simulation study of the annual maxima and the peaks-over-threshold methods Technical Report, SBW-Belastingen: Subproject "Statistics". Deltares Report 1200264-002.

Capéraà, P., Fougères, A. L. and Genest, C. (1997). A nonparametric estimation procedure for bivariate extreme value copulas. *Biometrika* **84** 567–577. . MR1603985 (99c:62078)

Coles, S. and Tawn, J. (1996). Modelling extremes of the areal rainfall process. *Journal of the Royal Statistical Society. Series B.* **58** 329–347.

Danielsson, J., de Haan, L., Peng, L. and de Vries, C. G. (2001). Using a bootstrap method to choose the sample fraction in tail index estimation. *Journal of Multivariate analysis* **76** 226–248.

De Haan, L. (1984). A Spectral Representation for Max-stable Processes. *The Annals of Probability* **12** 1194–1204.

de Haan, L. and Ferreira, A. (2006). *Extreme value theory: an introduction*. Springer.

de Haan, L. and Lin, T. (2003). Weak consistency of extreme value estimators in $C[0,1]$. *Annals of Statistics* **31** 1996–2012.

De Haan, L., Mercadier, C. and Zhou, C. (2016). Adapting extreme value statistics to financial time series: dealing with bias and serial dependence. *Finance and Stochastics* **20** 321–354.

de Haan, L., Neves, C. and Peng, L. (2008). Parametric tail copula estimation and model testing. *J. Multivariate Anal.* **99** 1260–1275. . MR2419346

de Haan, L. and Resnick, S. I. (1977). Limit theory for multivariate sample extremes. *Z. Wahrscheinlichkeitstheorie und Verw. Gebiete* **40** 317–337. . MR0478290

de Haan, L. and Stadtmüller, U. (1996). Generalized regular variation of second order. *J. Austral. Math. Soc. Ser. A* **61** 381–395. MR1420345

Dieker, A. and Mikosch, T. (2015). Exact simulation of Brown-Resnick random fields at a finite number of locations. *Extremes* **18** 301–314.

Dombry, C. (2015). Existence and consistency of the maximum likelihood estimators for the extreme value index within the block maxima framework. *Bernoulli* **21** 420–436.

Dombry, C., Engelke, S. and Oesting, M. (2016). Asymptotic properties of the maximum likelihood estimator for multivariate extreme value distributions. *ArXiv e-prints*.

Dombry, C., Engelke, S. and Oesting, M. (2016). Exact simulation of max-stable processes. *Biometrika* **103** 303-317.

Dombry, C., Éyi-Minko, F. and Ribatet, M. (2013). Conditional simulation of max-stable processes. *Biometrika* **100** 111–124.

Dombry, C. and Ferreira, A. (2017). Maximum likelihood estimators based on the block maxima method. ArXiv e-prints, arXiv:1705.00465.

Drees, H. (1998). On smooth statistical tail functionals. *Scand. J. Statist.* **25** 187–210. . MR1614276

Drees, H. (2000). Weighted approximations of tail processes for β-mixing random variables. *Ann. Appl. Probab.* **10** 1274–1301. . MR1810875

Drees, H., de Haan, L. and Resnick, S. (2000). How to make a Hill plot. *Ann. Statist.* **28** 254–274. . MR1762911

Drees, H., de Haan, L. and Li, D. (2003). On large deviation for extremes. *Statist. Probab. Lett.* **64** 51–62. . MR1995809

Drees, H. and de Haan, L. (2015). Estimating failure probabilities. *Bernoulli* **21** 957–1001. . MR3338653

Drees, H., Ferreira, A. and de Haan, L. (2004). On maximum likelihood estimation of the extreme value index. *Ann. Appl. Probab.* **14** 1179–1201.

Drees, H. and Huang, X. (1998). Best attainable rates of convergence for estimates of the stable tail dependence functions. *J. Multivar. Anal.* **64** 25-47.

Drees, H. and Kaufmann, E. (1998). Selecting the optimal sample fraction in univariate extreme value estimation. *Stochastic Processes and their Applications* **75** 149–172.

Drees, H. and Rootzén, H. (2010). Limit theorems for empirical processes of cluster functionals. *Ann. Statist.* **38** 2145–2186. . MR2676886

Einmahl, J. H. J., de Haan, L. and Piterbarg, V. I. (2001). Nonparametric estimation of the spectral measure of an extreme value distribution. *Ann. Statist.* **29** 1401–1423. . MR1873336

Einmahl, J. H. J., Krajina, A. and Segers, J. (2012). An M-estimator for tail dependence in arbitrary dimensions. *Ann. Statist.* **40** 1764–1793. . MR3015043

Einmahl, J. H. and Lin, T. (2006). Asymptotic normality of extreme value estimators on C [0, 1]. *The Annals of Statistics* **34** 469–492.

Einmahl, J. H. J. and Segers, J. (2009). Maximum empirical likelihood estimation of the spectral measure of an extreme-value distribution. *Ann. Statist.* **37** 2953–2989. . MR2541452

Ferreira, A. and De Haan, L. (2014). The generalized Pareto process; with a view towards application and simulation. *Bernoulli* **20** 1717–1737.

Ferreira, A. and de Haan, L. (2015). On the block maxima method in extreme value theory: PWM estimators. *Ann. Statist.* **43** 276–298.

Ferro, C. A. T. and Segers, J. (2003). Inference for clusters of extreme values. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **65** 545–556.

Fougères, A.-L., De Haan, L., Mercadier, C. et al. (2015). Bias correction in multivariate extremes. *The Annals of Statistics* **43** 903–934.

Genest, C., Ghoudi, K. and Rivest, L. P. (1995). A semiparametric estimation procedure of dependence parameters in multivariate families of distributions. *Biometrika* **82** 543–552. . MR1366280 (96j:62081)

Genest, C. and Segers, J. (2009). Rank-based inference for bivariate extreme-value copulas. *Ann. Statist.* **37** 2990–3022. . MR2541453 (2010i:62095)

Genest, C. and Segers, J. (2010). On the covariance of the asymptotic empirical copula process. *J. Multivariate Anal.* **101** 1837–1845. . MR2651959 (2011j:60107)

Giné, E., Hahn, M. G. and Vatan, P. (1990). Max-infinitely divisible and max-stable sample continuous processes. *Probability theory and related fields* **87** 139–165.

Gomes, M. I., De Haan, L. and Rodrigues, L. H. (2008). Tail index estimation for heavy-tailed models: accommodation of bias in weighted log-excesses. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **70** 31–52.

Gudendorf, G. and Segers, J. (2012). Nonparametric estimation of multivariate extreme-value copulas. *J. Statist. Plann. Inference* **142** 3073–3085. . MR2956794

Guillou, A. and Hall, P. (2001). A diagnostic for selecting the threshold in extreme value analysis. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **63** 293–305.

Gumbel, E. J. (1958). *Statistics of extremes*. Columbia University Press, New York. MR0096342 (20 ##2826)

Hall, P. and Welsh, A. H. (1984). Best attainable rates of convergence for estimates of parameters of regular variation. *The Annals of Statistics* **12** 1079–1084.

Hill, B. M. (1975). A Simple General Approach to Inference about the Tail of a Distribution. *Ann. Statist.* **3** 1163–1174.

Hosking, J. R. M., Wallis, J. R. and Wood, E. F. (1985). Estimation of the generalized extreme-value distribution by the method of probability-weighted moments. *Technometrics* **27** 251–261. . MR797563

Hsing, T. (1991). On Tail Index Estimation Using Dependent Data. *Ann. Statist.* **19** 1547–1569.

Hsing, T. (1993). Extremal index estimation for a weakly dependent stationary sequence. *Ann. Statist.* **21** 2043–2071.

Huang, X. (1992). Statistics of bivariate extreme values. PhD thesis, Tinbergen Institute Research Series, Netherlands.

Huser, R. and Davison, A. (2014). Space-time modelling of extreme events. *Journal of the Royal Statistical Society. Series B.* **76** 439–461.

Huser, R. G. and Wadsworth, J. L. (2017). Modeling spatial processes with unknown extremal dependence class. *Journal of the American Statistical Association* Forthcoming.

Kabluchko, Z., Schlather, M. and De Haan, L. (2009). Stationary max-stable fields associated to negative definite functions. *The Annals of Probability* **37** 2042–2065.

Leadbetter, M. R. (1974). On extreme values in stationary sequences. *Z. Wahrscheinlichkeitstheorie und Verw. Gebiete* **28** 289–303. . MR0362465

Leadbetter, M. R. (1983). Extremes and local dependence in stationary sequences. *Z. Wahrsch. Verw. Gebiete* **65** 291–306. . MR722133 (85b:60033)

Nandagopalan, S. (1994). On the multivariate extremal index. *Journal of Research-National Institute of Standards and Technology* **99** 543–543.

Naveau, P., Guillou, A., Cooley, D. and Diebolt, J. (2009). Modelling pairwise dependence of maxima in space. *Biometrika* **96** 1–17. . MR2482131

Northrop, P. J. (2015). An efficient semiparametric maxima estimator of the extremal index. *Extremes* **18** 585–603.

Oesting, M., Schlather, M. and Zhou, C. (2018). Exact and fast simulation of max-stable processes on a compact set using the normalized spectral representation. *Bernoulli* **24** 1497–1530.

Padoan, S. A., Ribatet, M. and Sisson, S. A. (2010). Likelihood-based inference for max-stable processes. *J. Amer. Statist. Assoc.* **105** 263–277. . MR2757202

Pickands, J. (1975). Statistical Inference Using Extreme Order Statistics. *Ann. Statist.* **3** 119–131.

Pickands, J. III (1981). Multivariate extreme value distributions. In *Proceedings of the 43rd session of the International Statistical Institute, Vol. 2 (Buenos Aires, 1981)* **49** 859–878, 894–902. With a discussion. MR820979

Prescott, P. and Walden, A. T. (1980). Maximum likelihood estimation of the parameters of the generalized extreme-value distribution. *Biometrika* **67** 723–724. . MR601119 (81m:62046)

Resnick, S. I. (1987). *Extreme values, regular variation, and point processes. Applied Probability. A Series of the Applied Probability Trust* **4**. Springer-Verlag, New York. . MR900810

Resnick, S. and Stărică, C. (1998). Tail index estimation for dependent data. *Ann. Appl.*

Probab. **8** 1156–1183.

Robert, C. Y. (2009). Inference for the limiting cluster size distribution of extreme values. *Ann. Statist.* **37** 271–310.

Robert, C. Y., Segers, J. and Ferro, C. A. T. (2009). A sliding blocks estimator for the extremal index. *Electron. J. Stat.* **3** 993–1020. . MR2540849

Rootzén, H. (2009). Weak convergence of the tail empirical process for dependent sequences. *Stochastic Process. Appl.* **119** 468–490. . MR2494000

Schmidt, R. and Stadtmüller, U. (2006). Non-parametric estimation of tail dependence. *Scand. J. Statist.* **33** 307–335. . MR2279645

Smith, R. L. and Weissman, I. (1994). Estimating the extremal index. *J. Roy. Statist. Soc. Ser. B* **56** 515–528.

Süveges, M. (2007). Likelihood estimation of the extremal index. *Extremes* **10** 41–55.

Thibaud, E. and Opitz, T. (2015). Efficient inference and simulation for elliptical Pareto processes. *Biometrika* **102** 855–870.

Weissman, I. and Novak, S. Y. (1998). On blocks and runs estimators of the extremal index. *J. Statist. Plann. Inference* **66** 281–288.

Zhang, D., Wells, M. T. and Peng, L. (2008). Nonparametric estimation of the dependence function for a multivariate extreme value distribution. *J. Multivariate Anal.* **99** 577–588. . MR2406072 (2009e:62225)