REVSTAT – Statistical Journal
Volume 6, Number 1, March 2008

IMPROVING PROBABILITY-WEIGHTED MOMENT METHODS FOR THE GENERALIZED EXTREME VALUE DISTRIBUTION

Authors: Jean Diebolt
– CNRS, Université de Marne-la-Vallée,
5, boulevard Descartes, 77454 Marne-la-Vallée Cedex 2, France

Armelle Guillou
– Université de Strasbourg, IRMA,
7, rue René Descartes, 67084 Strasbourg Cedex, France
guillou@math.u-strasbg.fr

Philippe Naveau
– Laboratoire des Sciences du Climat et de l'Environnement, IPSL-CNRS,
Orme des Merisiers / Bat. 709, 91191 Gif-sur-Yvette, France
philippe.naveau@lsce.ipsl.fr

Pierre Ribereau
– Université de Montpellier, Équipe de Probabilités et Statistique,
CC 051, Place Eugène Bataillon, 34095 Montpellier Cedex 5, France
pribere@math.univ-montp2.fr

Abstract:

- In 1985 Hosking *et al.* estimated with the so-called Probability-Weighted Moments (PWM) method the parameters of the Generalized Extreme Value (GEV) distribution, the latter being classically fitted to maxima of sequences of independent and identically distributed random variables. Their approach is still very popular in hydrology and climatology because of its conceptual simplicity, its easy implementation and its good performance for most distributions encountered in geosciences. Its main drawback resides in its limitations when applied to strong heavy-tailed densities. Whenever the GEV shape parameter is larger than 0.5, the asymptotic properties of the PWMs cannot be derived and consequently, asymptotic confidence intervals cannot be obtained. To broaden the validity domain of the PWM approach, we take advantage of a recent extension of PWM to a larger class of moments, called Generalized PWM (GPWM). This allows us to derive the asymptotic properties of our estimators for larger values of the shape parameter. The performance of our approach is illustrated by studying simulations of small, medium and large GEV samples. Comparisons with other GEV estimation techniques used in hydrology and climatology are performed.

Key-Words:

- *empirical processes; maximum likelihood estimators.*

AMS Subject Classification:

- 62G32, 60G70, 62G20.

1. INTRODUCTION

In climatology and hydrology, maxima of temperatures, precipitation and river discharges have been recorded for many decades. The block maxima size (hourly, daily, weekly, monthly or yearly) varies according to instrumental constraints, seasonalities and the application at hand. Extreme Value Theory (EVT) provides a theoretical framework to model the distribution of such block maxima (e.g. Embrechts *et al.*, 1997; Beirlant *et al.*, 2004; de Haan and Ferreira, 2006). Since the work of Fisher and Tippett in 1928, it is known that the only possible limiting form of a normalized maximum of a random sample (when a non-degenerate limit exists) is captured by the Generalized Extreme Value distribution (GEV)

$$G(x;\sigma,\gamma,\mu) = \begin{cases} \exp\left(-\left\{1+\gamma\,\dfrac{x-\mu}{\sigma}\right\}^{-1/\gamma}\right), & \text{if } 1+\gamma\,\dfrac{x-\mu}{\sigma}>0,\ \ \gamma\neq 0\ , \\ \exp\left(-\exp\left\{-\,\dfrac{x-\mu}{\sigma}\right\}\right), & \text{if } x\in\mathbb{R},\ \ \gamma=0\ , \end{cases}$$

with $\mu \in \mathbb{R}$, $\sigma > 0$ and γ are called the location, scale and shape parameters, respectively.

Whenever all observations from a given sample are available, it is statistically more efficient to disregard the block maxima modeling approach and instead to analyze exceedances above a high fixed threshold. The exceedances amplitudes can be asymptotically modeled by the Generalized Pareto Distribution (GPD) (e.g., Pickands, 1975; Davison, 1984). In the last four decades, a wide range of methods have been proposed to estimate the GPD scale and shape parameters (e.g. Embrechts *et al.*, 1997; Beirlant *et al.*, 2004; de Haan and Ferreira, 2006). But, for some specific cases, a GEV based approach may still be preferred to a GPD one for at least three reasons. Firstly, block maxima may be the only measurements available to the practitioner (this is specially true for long historical records). Secondly, climatologists frequently face a computational problem. A very high number time series have to be analyzed. For example, General Circulation Models, complex computer codes simulating the atmospheric circulation through resolving the equations representing the Earths atmospheric dynamics provide synthetic temperature time series on a spherical grid. The number of points on such a grid can easily be greater than the hundreds. Consequently, it is computationally easier to only focus on block maxima. This strategy bypasses the difficult problem of choosing a high threshold for each grid point (Kharin, 2007). The latter task is already difficult for a single time series. The third reason to work with blocks of a given size centers on the interpretability of the estimated parameters. For example, a block size of one year makes sense for the Earth scientist because inter-annual physical processes are often very different than decadal ones. For these three reasons, modeling block maxima with a GEV distribution remains a very frequent procedure in hydrology and climatology.

To estimate the GEV parameters in the independent and identically distributed (iid) setting, there exists a wide variety of approaches. In this paper we focus on the two most popular ones used in hydrology and climatology: method-of-moments types (e.g. Hosking *et al.*, 1985) and likelihood based procedures (e.g. Coles and Dixon, 1999; Katz *et al.*, 2002). For the former, hydrologists frequently analyze their maxima with the so-called Probability Weighted Moments (PWM) method introduced by Landwehr *et al.* (1979) and Greenwood *et al.* (1979). The main idea of this approach is to match the moments

$$\mathbb{E}\Big[X^p\big(F(X)\big)^r\big(1-F(X)\big)^s\Big], \qquad \text{with } p,\ r \text{ and } s \text{ real numbers },$$

with their empirical functionals, similarly to the classical method-of-moments. For the GEV distribution, it is easy to show (Hosking *et al.*, 1985) that $\mathbb{E}[X(F(X))^r]$ can be written as

$$\beta_r = \frac{1}{r+1}\left\{\mu - \frac{\sigma}{\gamma}\Big[1-(r+1)^\gamma\,\Gamma(1-\gamma)\Big]\right\}, \qquad \gamma<1 \quad\text{and}\quad \gamma\neq 0 . \tag{1.1}$$

Consequently, the PWM estimators $(\widehat{\sigma},\widehat{\gamma},\widehat{\mu})$ of the GEV parameters (σ,γ,μ) are simply the solution of the following system of equations

$$\begin{cases} \beta_0 \;=\; \mu - \dfrac{\sigma}{\gamma}\Big(1-\Gamma(1-\gamma)\Big) \\ 2\beta_1-\beta_0 \;=\; \dfrac{\sigma}{\gamma}\,\Gamma(1-\gamma)\,(2^\gamma-1) \\ \dfrac{3\beta_2-\beta_0}{2\beta_1-\beta_0} \;=\; \dfrac{3^\gamma-1}{2^\gamma-1} \end{cases}$$

in which β_r has to be replaced by the unbiased estimator proposed by Landwehr *et al.* (1979)

$$\widehat{\beta}_r \;=\; \frac{1}{n}\sum_{j=1}^{n}\left(\prod_{\ell=1}^{r}\frac{j-\ell}{n-\ell}\right)X_{j,n}$$

where $(X_{1,n},...,X_{n,n})$ represents the ordered GEV distributed sample. The properties and performances of $(\widehat{\beta}_0,\widehat{\beta}_1,\widehat{\beta}_2)$ and $(\widehat{\sigma},\widehat{\gamma},\widehat{\mu})$ were studied in details by Hosking *et al.* (1985) who showed the asymptotic normality of these estimators for $\gamma<0.5$. Hoskings and his co-workers also asserted that PWMs estimators performed better than a classical maximum likelihood estimation (MLE) for small samples (see also Hosking and Wallis, 1987). Its conceptual simplicity, its practicability and its good properties for small samples can explain the success of the PWM approach in geosciences (e.g. Katz *et al.*, 2002). Furrer and Naveau (2007) derived some PWMs properties for small GPD distributed samples.

Despite its qualities, the PWM approach has been criticized by Coles and Dixon (1999). In particular, these authors first argued that the PWM estimator assumes *a priori* that the GEV shape parameter is smaller than one, equivalent to specifying that the studied distribution has finite mean. Then they deduced that, if this prior information is available, then a penalized likelihood approach with the constraint $\gamma < 1$ should be preferred. In this case, a simulation study indicated that the penalized MLE outperformed the PWM estimators. But one has to be careful with such a reasoning because PWM estimators are still computable even when $\gamma > 1$ (like the sample mean $\overline{X}$ can be calculated even when the mean is not finite). A penalized MLE with the constraint $\gamma < 1$ will never be able to provide a shape estimator greater than one. In addition, the classical and penalized MLE approaches impose a restriction on the lower values of γ. We need $\gamma > -0.5$ to have regularity of the MLE based estimators and the numerical solutions of the MLE equations are erratic for γ close to -0.5. Although it is rare to work with bounded upper tails, they can be encountered in geophysics. For example, atmospheric scientists can be interested in relative humidity maxima, a bounded random variable. In this context, we argue that it is always better to try removing restrictions on γ than adding ones because we never know in practice the true value of the shape parameter. Hence one of our goals is to extend the validity of method-of-moments based procedures. Still, we agree with Coles and Dixon (1999) on the inherent flexibility of the maximum likelihood and that the conditions on moments existence have to be carefully examined and discussed to understand the limits of the PWMs approach. Our main point is not to sell one estimator in favor of another, but rather to know how to improve a simple approach frequently used in geosciences. With this objective in mind, we recall that Diebolt *et al.* (2007) have recently proposed a wider class of PWMs (called Generalized PWMs) for the GPD. In this paper, our aims are threefold. Firstly, we propose GPWM estimators for the GEV parameters. Secondly, we establish the asymptotic properties of our new estimators under general conditions ensuring the validity of the method for a large range of values of γ. Thirdly, we compare their performances with MLE and classical PWMs.

2. ASYMPTOTIC PROPERTIES OF THE GENERALIZED PWM ESTIMATORS

The generalized probability-weighted moments (GPWM) recently introduced by Diebolt *et al.* (2007) can be described in the following way

$$\nu_\omega = \mathbb{E}\big(X\,\omega(G)\big) = \int_{-\infty}^{\infty} x\,\omega\big(G(x)\big)\,dG(x)\ ,$$

where ω is a suitable continuous function. By changing variables, this moment can be rewritten as

$$\nu_\omega = \int_0^1 G^{-1}(u)\,\omega(u)\,du\ .$$

Let W be the primitive of ω, null at 0, i.e. $W(t) = \int_0^t \omega(u)\,du$. We propose to estimate ν_ω by

$$\widehat{\nu}_{\omega,n} = \int_0^1 \mathbb{F}_n^{-1}(u)\,\omega(u)\,du \tag{2.1}$$

where $\mathbb{F}_n$ denotes the classical empirical distribution function based on a sample $(X_1, ..., X_n)$. We are interested in the asymptotic properties of $\widehat{\nu}_{\omega,n}$ for the GEV distribution. To reach this goal, we select a function ω such that

$$\omega(t) = O\big((1-t)^b\big) \qquad \text{for } t \text{ close to } 1, \;\; b \geq 0 \tag{2.2}$$

and

$$\omega(t) = O(t^{a'}) \qquad \text{for } t \text{ close to } 0, \;\; a' > 0\;. \tag{2.3}$$

These assumptions tie down the functions $G^{-1}(t)$ and $\mathbb{F}_n^{-1}(t)$ at $t=0$ and $t=1$. An example of such a function is $\omega(t) = t^a(-\log t)^b$, $a > a'$. In this case, the GPWM for the GEV distribution can be rewritten (see Appendix) as

$$\nu_\omega = \frac{\sigma}{\gamma}\,\frac{1}{(a+1)^{b-\gamma+1}}\;\Gamma\big(b-\gamma+1\big) - \Big(\frac{\sigma}{\gamma}-\mu\Big)\,\frac{1}{(a+1)^{b+1}}\;\Gamma\big(b+1\big)\;. \tag{2.4}$$

Compared to Equality (1.1) derived by Hosking *et al.* (1985), we have a more general expression, Equation (1.1) can be obtained by taking $b=0$ in (2.4). As for the PWMs method, a system of three equations for three different values of a and/or b has to be solved in order to obtain estimators for σ, γ, μ.

Under the conditions (2.2) and (2.3), the GPWM ν_ω exists as soon as $\gamma < b+1$. This means that the domain of validity for the asymptotic normality of the GPWM estimators has been extended from the set $(\gamma < 1/2)$ to the larger set $\gamma < \frac{1}{2} + b$. More precisely, the following theorem summarizes our findings.

Theorem 2.1. *Let $(X_1, ..., X_n)$ be a sample of maxima whose marginal follows a GEV distribution. Let ω_1, ω_2 and ω_3 be any three continuous functions satisfying (2.2) and (2.3). If $\gamma < \frac{1}{2} + \min(b_1, b_2, b_3)$ for some $b_i \geq 0$, then the rescaled trivariate GPWM estimator vector defined by (2.1) and denoted by*

$$\sqrt{n}\begin{pmatrix} \widehat{\nu}_{\omega_1,n} - \nu_{\omega_1} \\ \widehat{\nu}_{\omega_2,n} - \nu_{\omega_2} \\ \widehat{\nu}_{\omega_3,n} - \nu_{\omega_3} \end{pmatrix}$$

converges in distribution towards the trivariate vector

$$\begin{pmatrix} \sigma \displaystyle\int_0^1 \frac{B(t)}{t}\,\big(-\log t\big)^{-\gamma-1}\omega_1(t)\,dt \\ \sigma \displaystyle\int_0^1 \frac{B(t)}{t}\,\big(-\log t\big)^{-\gamma-1}\omega_2(t)\,dt \\ \sigma \displaystyle\int_0^1 \frac{B(t)}{t}\,\big(-\log t\big)^{-\gamma-1}\omega_3(t)\,dt \end{pmatrix} \tag{2.5}$$

where B denotes a Brownian bridge and $n \to \infty$. The elements of the variance-covariance matrix, Γ, of this limiting vector are given by

$$
\begin{aligned}
(2.6)\qquad & \int_0^1 \frac{1}{t}\left(-\log t\right)^{-\gamma-1} \omega_i(t) \int_0^t \left(-\log s\right)^{-\gamma-1} \omega_j(s)\; ds\, dt \;\; + \\
& + \int_0^1 \left(-\log t\right)^{-\gamma-1} \omega_i(t) \int_0^t \frac{1}{s}\left(-\log s\right)^{-\gamma-1} \omega_j(s)\; ds\, dt \\
& - \int_0^1 \left(-\log t\right)^{-\gamma-1} \omega_i(t)\; dt \int_0^1 \left(-\log t\right)^{-\gamma-1} \omega_j(t)\; dt\;,
\end{aligned}
$$

where $i = 1,2,3$ *and* $j = 1,2,3$.

The proof of this theorem is postponed to the appendix and is based on empirical process arguments. From this result, we can deduce estimators for the three parameters σ, γ, μ of the GEV distribution by applying the delta-method. In order to assess the performance of our approach, we analyze simulated and real data in the next section.

3. ANALYSIS OF SIMULATED AND REAL DATA

Theorem 2.1 is a general result. In practice, we have to select the three function ω_1, ω_2 and ω_3. In this section, we opt for $\omega(t) = t^a(-\log t)^b$ with the three pairs $(a,b) = (1,1),(1,2),(2,1)$. This choice is justified by the fact that an estimator of γ can be deduced for these functions by solving the following equation

$$
\frac{\widehat{\gamma}}{1-\left(\frac{3}{2}\right)^{\widehat{\gamma}}} \;=\; \frac{2\left[\widehat{\omega}_{11} - \widehat{\omega}_{12}\right]}{\widehat{\omega}_{11} - \frac{9}{4}\,\widehat{\omega}_{21}}\;,
$$

where

$$
\widehat{\omega}_{ab} = \int_0^1 \mathbb{F}_n^{-1}(u)\, u^a (-\log u)^b\, du\;.
$$

Two estimators of σ and μ can be obtained from the relations

$$
\widehat{\sigma} \;=\; 2^{3-\widehat{\gamma}}\, \frac{\widehat{\omega}_{11} - \widehat{\omega}_{12}}{\Gamma(2-\widehat{\gamma})} \qquad \text{and} \qquad \widehat{\mu} \;=\; \frac{\widehat{\sigma}}{\widehat{\gamma}} - \frac{\widehat{\sigma}}{\widehat{\gamma}}\, 2^{\widehat{\gamma}}\, \Gamma(2-\widehat{\gamma}) + 4\,\widehat{\omega}_{11}\;.
$$

From Theorem 2.1, the asymptotic normality of these three estimators of the GEV parameters can be derived. It is possible to show the existence of a C^1-diffeomorphism T which transforms the GPWMs $(\omega_{11}, \omega_{12}, \omega_{21})$ into (σ, γ, μ). Direct but lengthy computations lead to the following Jacobian matrix, M, associated to this diffeomorphism

$$\begin{pmatrix} \dfrac{2^{\gamma-2}}{\gamma}\,\Gamma(2-\gamma)-\dfrac{1}{4\gamma} & \dfrac{\sigma}{\gamma}\,2^{\gamma-2}\Big[\Big(\log 2-\dfrac{1}{\gamma}\Big)\,\Gamma(2-\gamma)-\Gamma'(2-\gamma)\Big]+\dfrac{\sigma}{4\gamma^2} & \dfrac{1}{4} \\ \Big(\dfrac{2}{\gamma}-1\Big)\,2^{\gamma-3}\,\Gamma(2-\gamma)-\dfrac{1}{4\gamma} & \dfrac{\sigma}{\gamma}\,2^{\gamma-3}\Big[\Big(-\dfrac{2}{\gamma}+\log 2\,(2-\gamma)\Big)\,\Gamma(2-\gamma)-(2-\gamma)\,\Gamma'(2-\gamma)\Big]+\dfrac{\sigma}{4\gamma^2} & \dfrac{1}{4} \\ \dfrac{3^{\gamma-2}}{\gamma}\,\Gamma(2-\gamma)-\dfrac{1}{9\gamma} & \dfrac{\sigma}{\gamma}\,3^{\gamma-2}\Big[\Big(\log 3-\dfrac{1}{\gamma}\Big)\,\Gamma(2-\gamma)-\Gamma'(2-\gamma)\Big]+\dfrac{\sigma}{9\gamma^2} & \dfrac{1}{9} \end{pmatrix}.$$

Under the same assumptions stated in Theorem 2.1, we can deduce that the limiting variance-covariance matrix of the trivariate vector $\sqrt{n}\begin{pmatrix}\widehat{\sigma}-\sigma\\ \widehat{\gamma}-\gamma\\ \widehat{\mu}-\mu\end{pmatrix}$ can be written as $M^{-1}\Gamma(M^{-1})'$. This matrix will be useful for computing asymptotic confidence intervals for our GPWM estimators for our application.

3.1. A simulation study

The aim of this simulation study is to show that our method performs adequately for a wide range of values of γ (we will test $\gamma=-0.2$, 0, 0.2 and 1.2) and for small and medium samples sizes ($n=15$, 25, 50 and 100). The quality of our estimators will be compared to the two most common approaches used in hydrology (MLE and PWM). These three estimation methods (MLE, PWM and GPWM) are invariant under linear transformations of the data, so without loss of generality the location and scale parameters are set to $\mu=0$ and $\sigma=1$ in all the simulations. For each combination of values of n and γ, 10 000 random samples are generated from the GEV distribution, and for each sample of parameters, μ, σ and γ were estimated by each of the three methods.

We also implemented the penalized likelihood procedure proposed by Coles and Dixon (1999) but it did not produced valuable results for $\gamma=-0.2$ and 1.2 for all sample sizes. Consequently, we will not show figures about the penalized likelihood procedure (they are available upon request).

Figure 1 shows the estimation results for four different shape parameters γ. Each vertical panel corresponds to the estimations obtained from $\gamma=-0.2$, 0, 0.2 and 1.2 (from bottom to top). The gray, yellow and white boxplots derived from 10 000 GEV samples represent the performance of the MLE, PWM and GPWM estimators, respectively. The x-axis corresponds to different sample sizes $n=15$, 25, 50 and 100. This graph indicates at least three things for the estimation of the shape parameter. For $\gamma=-0.2$, 0, 0.2, the GPWM and MLE behave fairly similarly for all sample sizes, while the PWM method tends to a smaller interquartile but a larger bias. For strong heavy tail ($\gamma=1.2$), PWM does not perform well. The MLE provides a very large interquartile (even for $n=100$),

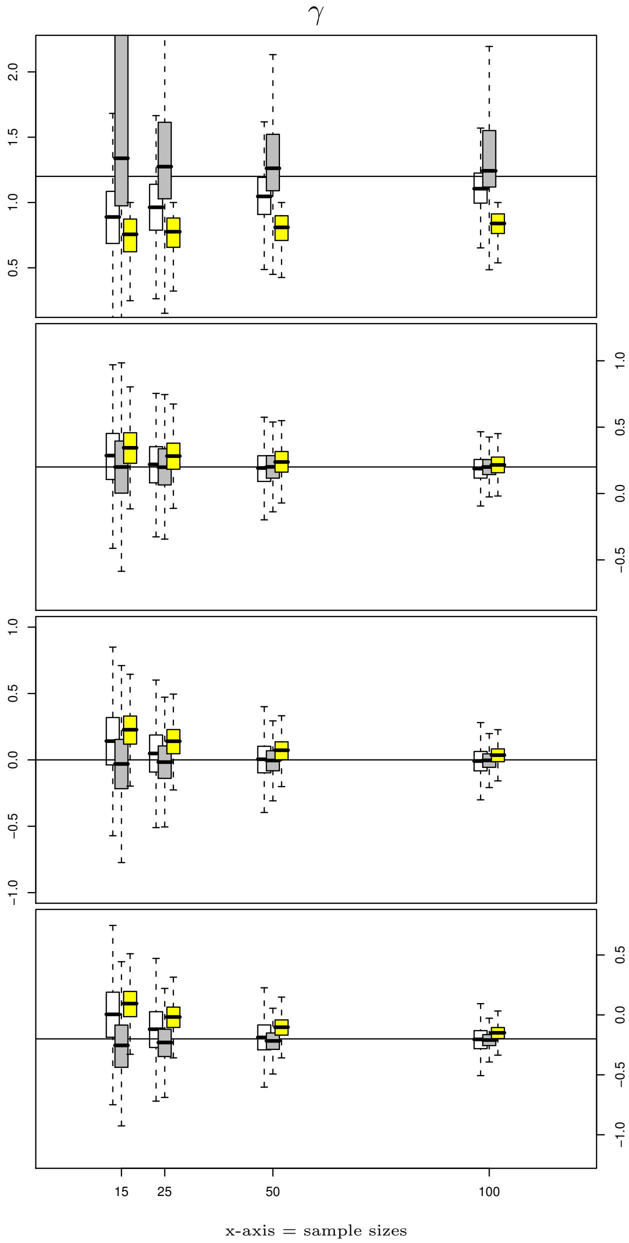

Figure 1: Estimation of γ: The gray, yellow and white boxplots from 10 000 GEV samples represent the performance of the MLE, PWM and GPWM estimators, respectively. The x-axis corresponds to different sample sizes $n = 15$, 25, 50 and 100. Each vertical panel represents the estimations obtained from a different value of $\gamma = -0.2$, 0, 0.2 and 1.2 (from bottom to top).

while the GPWM gives reasonable results for medium sample sizes $n = 50$ and $n = 100$. But this may not be the whole story because one also has to look at the two other GEV parameters. Figure 2 displays the estimation results for $\mu = 0$ (left panels) and $\sigma = 1$ (right panels). As in Figure 1, each vertical panel represents the estimations obtained from $\gamma = -0.2$, 0, 0.2 and 1.2 (from bottom to top).

Figure 2 confirms the remarks raised from Figure 1. The GPWM estimators seem to outperform the PWM ones in all cases. That means that our generalization of the PWM has widened the domain of validity without deteriorating the estimation of the parameters. The MLE approach works adequately but not for $\gamma = 1.2$. For this latter case, the estimation of σ even for $n = 100$ does not seem to provide reasonable estimates.

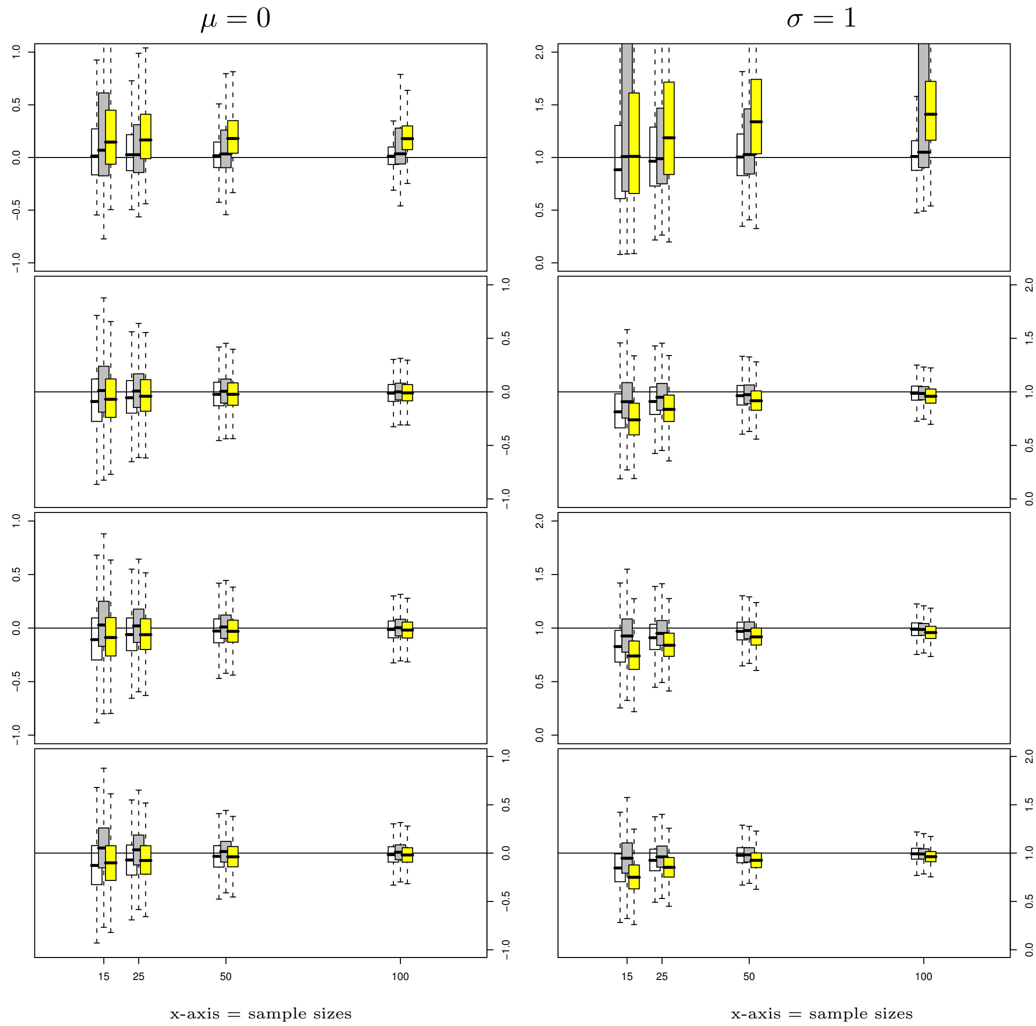

Figure 2: Estimation of $\mu = 0$ (left panels) and $\sigma = 1$ (right panels): The gray, yellow and white boxplots from 10 000 GEV samples represent the performance of the MLE, PWM and GPWM estimators, respectively. The x-axis corresponds to different sample sizes n = 15, 25, 50 and 100. Each vertical panel represents the estimations obtained from a different value of $\gamma = -0.2$, 0, 0.2 and 1.2 (from bottom to top).

3.2. A real data set

One weather station in the city of Fort-Collins (Colorado, USA) recorded annual daily precipitation maxima (in mm) from 1948 to 2001. Figure 3 displays these precipitation maxima. The year 1997 stands up because a storm caused extensive flood damage to this city on July 28$^{\text{th}}$ 1997. In order to fit a GEV distribution to this series of yearly maxima, we apply the three estimation methods

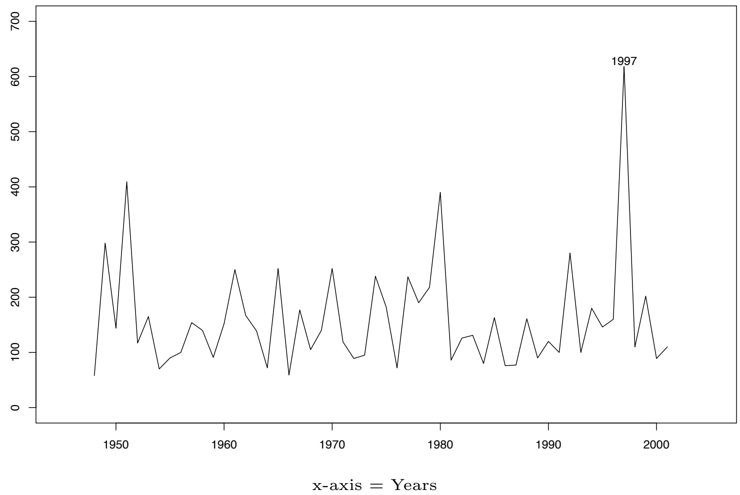

Figure 3: Annual daily precipitation maxima (in mm) recorded in Fort Collins (Colorado, USA) from 1948 to 2001.

(MLE, PWM and GPWM) to our data. For each method and for each parameter, 95% asymptotic confidence intervals were obtained. Table 1 summarizes our findings. The three estimation methods (PWM, MLE and GPWM) give similar value for the shape parameter γ, around 0.3. For this type of value and type of sample size (around 50), we know from our simulation study that the three methods should provide similar results in terms of estimation and confidence intervals. The results presented in Table 1 tends to confirm this fact.

Table 1: GEV parameters fitted to the annual daily precipitation maxima in Fort-Collins, Colorado, USA. For each estimation method and parameters, the 95% asymptotic confidence intervals are shown into brackets.

	$\hat{\mu}$	$\hat{\sigma}$	$\hat{\gamma}$
PWM	112.47 $[96.61, 128.33]$	50.57 $[36.43, 64.71]$	0.27 $[-0.01, 0.55]$
MLE	111.31 $[96.45, 126.17]$	47.39 $[34.43, 60.35]$	0.35 $[0.07, 0.63]$
GPWM	112.01 $[106.40, 117.62]$	50.24 $[38.12, 62.36]$	0.32 $[-0.08, 0.73]$

4. CONCLUDING REMARKS

In this paper, we extend the PWM method of Hosking *et al.* (1985) for the GEV distribution. As observed in the simulation part, the validity domain is not only broadened but also the performance of our new method is improved over the classical PWM, especially for large values of the shape parameter. The latter situation is not favorable to the ML approach for small and medium sample sizes. Still, while it is clear that GPWM should be favored to classical PWM, it is difficult to disregard the MLE because it can bring a powerful flexibility in the presence of covariates and/or non-stationarity. In the iid case, the hydrologist and the climatologist may prefer to estimate their GEV parameters with GPWMs because the latter are based on the same method-of-moment approach as the PWM. PWM has been used in their communities for decades and is well understood. The GPWM conserves the PWM conceptual simplicity and its easy implementation. Consequently, it could be quickly integrated in the toolbox of the hydrologist. One remaining challenge for the statistician is to extend such method-of-moment procedures to non-stationary situations.

APPENDIX

Proof of equality (2.4)

$$
\begin{aligned}
\nu_\omega &= \int_{-\infty}^{\infty} x\big(G(x)\big)^a \big(-\log G(x)\big)^b \, dG(x) \\
&= \int_0^1 G^{-1}(u)\, u^a \big(-\log u\big)^b \, du \\
&= \int_0^1 \left\{ \frac{\sigma}{\gamma} \Big[\big(-\log u\big)^{-\gamma} - 1 \Big] + \mu \right\} u^a \big(-\log u\big)^b \, du \\
&= \int_0^\infty \left\{ \frac{\sigma}{\gamma} \Big[x^{-\gamma} - 1 \Big] + \mu \right\} e^{-(a+1)x} x^b \, dx \\
&= \frac{\sigma}{\gamma} \int_0^\infty x^{b-\gamma} e^{-(a+1)x} \, dx \;-\; \left(\frac{\sigma}{\gamma} - \mu \right) \int_0^\infty x^b e^{-(a+1)x} \, dx \\
&= \frac{\sigma}{\gamma} \, \frac{1}{(a+1)^{b-\gamma+1}} \, \Gamma\big(b-\gamma+1\big) - \left(\frac{\sigma}{\gamma} - \mu \right) \frac{1}{(a+1)^{b+1}} \, \Gamma\big(b+1\big) \,.
\end{aligned}
$$

Proof of our Theorem 2.1

We consider the difference

$$
\begin{aligned}
\widehat{\nu}_{\omega,n} - \nu_\omega &\overset{d}{=} \int_0^1 \Big[G^{-1}\big(\mathbb{G}_n^{-1}(t)\big) - G^{-1}(t)\Big]\,\omega(t)\;dt \\
&= \int_0^{\frac{a_n}{n}} \Big[G^{-1}\big(\mathbb{G}_n^{-1}(t)\big) - G^{-1}(t)\Big]\,\omega(t)\;dt \\
&\quad + \int_{\frac{a_n}{n}}^{1-\frac{a_n}{n}} \Big[G^{-1}\big(\mathbb{G}_n^{-1}(t)\big) - G^{-1}(t)\Big]\,\omega(t)\;dt \\
&\quad + \int_{1-\frac{a_n}{n}}^{1} \Big[G^{-1}\big(\mathbb{G}_n^{-1}(t)\big) - G^{-1}(t)\Big]\,\omega(t)\;dt \\
&=: T_{1,n} + T_{2,n} + T_{3,n}\ ,
\end{aligned}
$$

where $(a_n)_n$ is defined by $a_n = ([9\log\log n] + 1)^2$ and $\mathbb{G}_n^{-1}$ denotes the empirical quantile function of independent uniform random variables on $(0,1)$. We study the different terms separately. We can easily prove that, if $\gamma \neq 0$, we have

$$
G^{-1}(t) \;=\; \frac{\sigma}{\gamma}\Big[\big(-\log t\big)^{-\gamma} - 1\Big] + \mu\ .
$$

The case $\gamma = 0$ can be viewed as the limiting case, letting $\gamma \to 0$.

<u>Term $T_{1,n}$</u>

$$
\begin{aligned}
T_{1,n} &= \int_0^{\frac{a_n}{n}} \frac{\sigma}{\gamma}\Big[\big(-\log \mathbb{G}_n^{-1}(t)\big)^{-\gamma} - \big(-\log t\big)^{-\gamma}\Big]\,\omega(t)\;dt \\
&= \frac{\sigma}{\gamma}\int_0^{\frac{a_n}{n}} \big(-\log \mathbb{G}_n^{-1}(t)\big)^{-\gamma}\omega(t)\;dt \;-\; \frac{\sigma}{\gamma}\int_0^{\frac{a_n}{n}} \big(-\log t\big)^{-\gamma}\omega(t)\;dt \\
&=: T_{1,n}^{(1)} + T_{1,n}^{(2)}\ .
\end{aligned}
$$

By changing variables, it is clear that

$$
T_{1,n}^{(2)} \;=\; -\frac{\sigma}{\gamma}\int_{\log\frac{n}{a_n}}^{\infty} x^{-\gamma}\,e^{-x}\,\omega(e^{-x})\;dx\ .
$$

Consequently

$$
\Big|T_{1,n}^{(2)}\Big| \;\le\; \frac{\sigma}{\gamma}\,\frac{a_n}{n}\int_{\log\frac{n}{a_n}}^{\infty} x^{-\gamma}\big|\omega(e^{-x})\big|\;dx\ .
$$

Therefore, we have, under the assumption

$$
\text{(A.1)} \qquad \int_0^{\infty} x^{-\gamma}\,\big|\omega(e^{-x})\big|\;dx \;<\; \infty\ ,
$$

that

$$\sqrt{n}\left|T_{1,n}^{(2)}\right| = O\left(\frac{a_n}{\sqrt{n}}\right) \longrightarrow 0 \ .$$

Of course, (A.1) is satisfied since we have (2.2) and (2.3). Now, concerning the term $T_{1,n}^{(1)}$, we use the following decomposition

$$\begin{aligned}
T_{1,n}^{(1)} &= \frac{\sigma}{\gamma}\left[\int_0^{1/n}\left(-\log \mathbb{G}_n^{-1}(t)\right)^{-\gamma}\omega(t)\,dt + \cdots + \int_{(a_n-1)/n}^{a_n/n}\left(-\log \mathbb{G}_n^{-1}(t)\right)^{-\gamma}\omega(t)\,dt\right] \\
&= \frac{\sigma}{\gamma}\sum_{i=1}^{a_n}\int_{\frac{i-1}{n}}^{\frac{i}{n}}\left(-\log \mathbb{G}_n^{-1}(t)\right)^{-\gamma}\omega(t)\,dt \\
&= \frac{\sigma}{\gamma}\sum_{i=1}^{a_n}\left(-\log U_{i,n}\right)^{-\gamma}\left[W\left(\frac{i}{n}\right) - W\left(\frac{i-1}{n}\right)\right] \\
&= \frac{\sigma}{\gamma}\,\frac{1}{n}\sum_{i=1}^{a_n}\left(-\log U_{i,n}\right)^{-\gamma}\omega\left(\frac{\xi_{i,n}}{n}\right)
\end{aligned}$$

with $i-1 \leq \xi_{i,n} \leq i$ and $U_{1,n} \leq ... \leq U_{n,n}$ the order statistics of a sample of n independent random variables from a uniform distribution on $(0,1)$. Since $|\omega(t)| \leq C\,t^{a'}$ for t close to 0, we have

$$\left|T_{1,n}^{(1)}\right| \ \leq\ \frac{\sigma}{\gamma}\,\frac{C}{n^{a'+1}}\left\{\left(-\log U_{1,n}\right)^{-\gamma} + \sum_{i=2}^{a_n}\left(-\log U_{i,n}\right)^{-\gamma} i^{a'}\right\} .$$

Using the following bounds (see Shorack and Wellner, 1986, p. 408 & 420), we have, for n large enough:

- for $U_{1,n}$:

$$\frac{1}{n(\log n)^{1+\varepsilon}} \ \leq\ U_{1,n} \ \leq\ (1+\varepsilon')\,\frac{\log\log n}{n} \qquad \text{a.s.}$$

- for $U_{i,n}$:

$$\max_{1\leq i\leq n}\frac{i}{n\,U_{i,n}} \ \leq\ \left(\log n\right)^{1+\varepsilon} \qquad \text{a.s.}\,,$$

$$\max_{1\leq i\leq n}\frac{n\,U_{i+1,n}}{i} \ \leq\ (1+\varepsilon')\log\log n \qquad \text{a.s.}\,.$$

Therefore, it is clear that

$$\sqrt{n}\left|T_{1,n}^{(1)}\right| = O\left(\frac{(\log n)^{-\gamma}}{n^{a'+1/2}}\right) + O\left(\frac{a_n^{a'+1}}{n^{a'+1/2}}\left(\log n\right)^{-\gamma}\right) \longrightarrow 0 \ ,$$

by definition of a_n.

Term $T_{2,n}$

$$\begin{aligned}
\sqrt{n}\,T_{2,n} &= \sqrt{n}\,\frac{\sigma}{\gamma}\int_{\frac{a_n}{n}}^{1-\frac{a_n}{n}} \left[\left(-\log\left(t+\frac{\beta_n(t)}{\sqrt{n}}\right)\right)^{-\gamma} - \left(-\log t\right)^{-\gamma}\right]\omega(t)\;dt\\
&= \sigma\int_{\frac{a_n}{n}}^{1-\frac{a_n}{n}} \frac{\beta_n(t)}{t}\left(-\log t\right)^{-\gamma-1}\omega(t)\;dt\\
&\quad + \frac{\sigma}{2\sqrt{n}}\int_{\frac{a_n}{n}}^{1-\frac{a_n}{n}} \frac{\beta_n^2(t)}{\xi_{t,n}^2}\left(-\log \xi_{t,n}\right)^{-\gamma-2}\left(\log\xi_{t,n}+\gamma+1\right)\omega(t)\;dt\;,
\end{aligned}$$

where β_n is the uniform empirical quantile process and $\xi_{t,n} \in \Big[\min\Big(t, t+\frac{\beta_n(t)}{\sqrt{n}}\Big), \max\Big(t, t+\frac{\beta_n(t)}{\sqrt{n}}\Big)\Big]$. Our aim now is to use a result due to Csörgő *et al.* (1983) (see e.g. Shorack and Wellner, 1986, p. 500). There exists a sequence of Brownian bridges B_n such that, for $\nu \in [0, \frac{1}{2}[$:

$$\begin{aligned}
\sqrt{n}\,T_{2,n} &= \sigma\, n^{-\nu}\int_{\frac{a_n}{n}}^{1-\frac{a_n}{n}} n^{\nu}\,\frac{\beta_n(t)-B_n(t)}{\big[t(1-t)\big]^{\frac{1}{2}-\nu}}\,\frac{1}{t}\left(-\log t\right)^{-\gamma-1}\big[t(1-t)\big]^{\frac{1}{2}-\nu}\omega(t)\;dt\\
&\quad + \sigma\int_{\frac{a_n}{n}}^{1-\frac{a_n}{n}} \frac{B_n(t)}{t}\left(-\log t\right)^{-\gamma-1}\omega(t)\;dt\\
&\quad + \frac{\sigma}{2\sqrt{n}}\int_{\frac{a_n}{n}}^{1-\frac{a_n}{n}} \frac{\beta_n^2(t)}{\xi_{t,n}^2}\left(-\log \xi_{t,n}\right)^{-\gamma-2}\left(\log\xi_{t,n}+\gamma+1\right)\omega(t)\;dt\\
&=:\; T_{2,n}^{(1)} + T_{2,n}^{(2)} + T_{2,n}^{(3)}\;,
\end{aligned}$$

with

$$\left|T_{2,n}^{(1)}\right| \;\le\; O_{\mathbb{P}}(n^{-\nu})\int_{\frac{a_n}{n}}^{1-\frac{a_n}{n}} \frac{1}{t}\left(-\log t\right)^{-\gamma-1}\big[t\,(1-t)\big]^{\frac{1}{2}-\nu}\omega(t)\;dt\;.$$

Therefore, under the conditions (2.2) and (2.3), $T_{2,n}^{(1)}$ tends to 0 as soon as $\gamma < b+\frac{1}{2}$. Now, we consider $T_{2,n}^{(2)}$. We can use the fact that

$$\int_{\frac{a_n}{n}}^{1-\frac{a_n}{n}} \frac{B_n(t)}{t}\left(-\log t\right)^{-\gamma-1}\omega(t)\;dt \;\overset{d}{=}\; \int_{\frac{a_n}{n}}^{1-\frac{a_n}{n}} \frac{B(t)}{t}\left(-\log t\right)^{-\gamma-1}\omega(t)\;dt$$

with

$$|B(t)| \;\le\; C\sqrt{\big[t\,(1-t)\big]\log\log\frac{1}{\big[t\,(1-t)\big]}} \qquad \text{a.s., for } t \text{ close to 0 and 1}\;.$$

Here and in all the paper, C represents a generic constant.

Therefore, we have

$$\int_{\frac{a_n}{n}}^{1-\frac{a_n}{n}} \frac{B_n(t)}{t}\left(-\log t\right)^{-\gamma-1}\omega(t)\;dt \;\xrightarrow{d}\; \int_0^1 \frac{B(t)}{t}\left(-\log t\right)^{-\gamma-1}\omega(t)\;dt\,, \qquad n\to\infty\,.$$

Now, we have to study $T^{(3)}_{2,n}$. According to Shorack and Wellner (1986, p. 616), we have

$$\big|\beta_n(t)\big| \le C\sqrt{t(1-t)}\,\sqrt{\log\log n} \quad \text{a.s., uniformly on } \left[9\,\frac{\log\log n}{n}, 1-9\,\frac{\log\log n}{n}\right].$$

Therefore

$$\left|T^{(3)}_{2,n}\right| \;\le\; C\,\frac{\log\log n}{\sqrt{n}} \int_{\frac{a_n}{n}}^{1-\frac{a_n}{n}} \frac{t\,(1-t)}{\xi^2_{t,n}}\,\big(-\log\xi_{t,n}\big)^{-\gamma-2}\big(\log\xi_{t,n}+\gamma+1\big)\,\omega(t)\;dt\;.$$

This integral can be divided into three parts: from $\frac{a_n}{n}$ to ε, from ε to $1-\varepsilon$ and from $1-\varepsilon$ to $1-\frac{a_n}{n}$, where ε is fixed. We denote these integrals by $T^{(3,1)}_{2,n}, T^{(3,2)}_{2,n}$ and $T^{(3,3)}_{2,n}$ respectively. We start with $T^{(3,1)}_{2,n}$. Note that for $t\in[\frac{a_n}{n},\varepsilon]$, we have

$$\text{(A.2)}\qquad \frac{\big|\beta_n(t)\big|}{t\,\sqrt{n}} \;\le\; C\sqrt{\frac{1-t}{t}}\,\sqrt{\frac{\log\log n}{n}}$$

$$\text{(A.3)}\qquad \le\; \frac{C}{\sqrt{\log\log n}}\;.$$

Therefore

$$\text{(A.4)}\qquad \frac{\xi_{t,n}}{t} = 1+o(1) \qquad \text{and} \qquad \log\xi_{t,n} = \big(1+o(1)\big)\,\log t\;,$$

where the $o(1)$-terms are uniform in t. Consequently,

$$\begin{aligned}\left|T^{(3,1)}_{2,n}\right| \;\le\; &-\,C\,\frac{\log\log n}{\sqrt{n}} \int_{\frac{a_n}{n}}^{\varepsilon} \frac{1-t}{t}\,\big(-\log t\big)^{-\gamma-1}\omega(t)\;dt\;\big(1+o(1)\big)\\ &+\;C\,(1+\gamma)\,\frac{\log\log n}{\sqrt{n}} \int_{\frac{a_n}{n}}^{\varepsilon} \frac{1-t}{t}\,\big(-\log t\big)^{-\gamma-2}\omega(t)\;dt\;\big(1+o(1)\big)\;.\end{aligned}$$

Using (2.3), $T^{(3,1)}_{2,n}$ tends clearly to 0. Similarly, since (A.2) and (A.4) are true for $t\in\big[1-\varepsilon, 1-\frac{a_n}{n}\big]$, we have

$$\left|T^{(3,3)}_{2,n}\right| \;=\; O\!\left(\frac{\log\log n}{\sqrt{n}}\right)\left\{\int_{1-\varepsilon}^{1-\frac{a_n}{n}}(1-t)^{b-\gamma}\,dt \;+\; \int_{1-\varepsilon}^{1-\frac{a_n}{n}}(1-t)^{b-\gamma-1}\,dt\right\},$$

by (2.2). The right-hand side of the last equality tends to 0 as soon as $\gamma < b+\frac{1}{2}$. For the central part, $T^{(3,2)}_{2,n}$, similar arguments lead to its negligibility.

<u>Term $T_{3,n}$</u>

$$\begin{aligned}\sqrt{n}\,T_{3,n} \;=\;& \frac{\sigma}{\gamma}\,\sqrt{n}\int_{1-\frac{a_n}{n}}^{1}\big(-\log\mathbb{G}_n^{-1}(t)\big)^{-\gamma}\omega(t)\;dt \;-\; \frac{\sigma}{\gamma}\,\sqrt{n}\int_{1-\frac{a_n}{n}}^{1}\big(-\log t\big)^{-\gamma}\omega(t)\;dt\\ =:&\; T^{(1)}_{3,n} + T^{(2)}_{3,n}\;.\end{aligned}$$

The term $T^{(2)}_{3,n}$ is of order

$$\sqrt{n}\int_{1-\frac{a_n}{n}}^{1}(1-t)^{-\gamma+b}\,dt$$

which tends to 0 as soon as $\gamma < b+\frac{1}{2}$. For $T^{(1)}_{3,n}$, we decompose again the integral as follows, with $j-1 \le \xi_j \le j$:

$$\begin{aligned}
T^{(1)}_{3,n} &= \frac{\sigma}{\gamma}\sqrt{n}\,\frac{1}{n}\sum_{j=n-a_n+1}^{n}\Big(-\log U_{j,n}\Big)^{-\gamma}\,\omega\Big(\frac{\xi_j}{n}\Big)\\
&= \frac{\sigma}{\gamma}\sqrt{n}\,\frac{1}{n}\sum_{j=1}^{a_n}\Big(1-U_{n-j+1,n}\Big)^{-\gamma}\left(1+O_{\mathbb{P}}\Big(\frac{j}{n}\Big)\right)\omega\Big(\frac{\xi_{n-j+1}}{n}\Big)\\
&\stackrel{d}{=} \frac{\sigma}{\gamma}\sqrt{n}\,\frac{1}{n}\sum_{j=1}^{a_n}\Big(U_{j,n}\Big)^{-\gamma}\left(1+O_{\mathbb{P}}\Big(\frac{j}{n}\Big)\right)\omega\Big(\frac{\xi_{n-j+1}}{n}\Big)\\
&= O\Big(\big(\log n\big)^{|\gamma|\,(1+\varepsilon)}\,n^{\gamma-b-\frac{1}{2}}\,a_n^{b-\gamma+1}\Big) \;=\; o(1)\;,
\end{aligned}$$

as soon as $\gamma < b+\frac{1}{2}$.

From all the above convergences, using Serfling (1980, page 18), we deduce (2.5), and therefore the expression of the generic term at position (i,j), $1 \le i,j \le 3$, of the limiting variance-covariance matrix given in (2.6).

Combining these results, Theorem 2.1 follows.

ACKNOWLEDGMENTS

This work was supported by the french ANR-AssimilEx project, the european E2-C2 grant, and by The Weather and Climate Impact Assessment Science Initiative at the National Center for Atmospheric Research (NCAR). The authors would also like to credit the contributors of the R project. Moreover, the authors are grateful to the referee for her/his constructive comments from which the paper has benefitted a lot.

REFERENCES

[1] Beirlant, J.; Goegebeur, Y.; Segers, J. and Teugels, J. (2004). *Statistics of Extremes: Theory and Applications*, Wiley Series in Probability and Statistics.

[2] Coles, S.G. and Dixon, M.J. (1999). Likelihood-based inference of extreme value models, *Extremes*, **2**, 5–23.

[3] Csörgő, M.; Csörgő, S.; Horváth, L. and Mason, D.M. (1983). An asymptotic theory for empirical reliability and concentration processes, *Unpublished manuscript*.

[4] Davison, A.C. (1984). *Modelling excesses over high thresholds, with an application*. In "Statistical Extremes and Applications" (J. Tiago de Oliveira, Ed.), Vol. 131, pp. 461–482, Dordrecht, D. Reidel Publishing Co..

[5] de Haan, L. and Ferreira, A. (2006). *Extreme Value Theory: An Introduction*, Springer Series in Operations Research.

[6] Diebolt, J.; Guillou, A. and Rached, I. (2007). Approximation of the distribution of excesses through a generalized probability-weighted moments method, *J. Statist Plann. Inference*, **137**, 841–857.

[7] Embrechts, P.; Klüppelberg, C. and Mikosch, T. (1997). *Modelling Extremal Events for Insurance and Finance*, Vol. 33 of "Applications of Mathematics", Springer-Verlag, Berlin.

[8] Fisher, R.A. and Tippett, L.H.C. (1928). Limiting forms of the frequency distribution in the largest particle size and smallest member of a sample, *Proc. Camb. Phil. Soc.*, **24**, 180–190.

[9] Furrer, R. and Naveau, P. (2007). Probability weighted moments properties for small samples, *Stat. Probab. Letters*, **70**, 190–195.

[10] Greenwood, J.A.; Landwehr, J.M.; Matalas, N.C. and Wallis, J.R. (1979). Probability-weighted moments: definition and relation to parameters of several distributions expressable in inverse form, *Water Resources Research*, **15**, 1049–1054.

[11] 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.

[12] Hosking, J.R.M. and Wallis, J.R. (1987). Parameter and quantile estimation for the generalized Pareto distribution, *Technometrics*, **29**, 339–349.

[13] Katz, R.; Parlange, M. and Naveau, P. (2002). Extremes in hydrology, *Advances in Water Resour.*, **25**, 1287–1304.

[14] Kharin, V.V.; Zwiers, F.W.; Zhang, X. and Hegerl, G.C. (2007). Changes in temperature and precipitation extremes in the IPCC ensemble of global coupled model simulations, *Journal of Climate*, **20**, 1419–1444.

[15] Landwehr, J.; Matalas, N. and Wallis, J. (1979). Probability weighted moments compared with some traditional techniques in estimating Gumbel parameters and quantiles, *Water Resour. Res.*, **15**, 1055–1064.

[16] Pickands, J. (1975). Statistical inference using extreme order statistics, *Ann. Statist.*, **3**, 119–131.

[17] Serfling, R.J. 1980. *Approximation Theorems of Mathematical Statistics*, Wiley & Son.

[18] Shorack, G.R. and Wellner, J.A. 1986. *Empirical Processes with Applications to Statistics*, Wiley, New York.

[19] Zhang J. (2007). Likelihood Moment Estimation for the Generalized Pareto Distribution, *Aust. N. Z. J. Stat.*, **49**, 69–77.