Modeling jointly low, moderate, and heavy rainfall intensities without a threshold selection

Philippe Naveau, Raphaël Huser, Pierre Ribereau, Alexis Hannart

AGU PUBLICATIONS

Water Resources Research

RESEARCH ARTICLE
10.1002/2015WR018552

Key Points:
- Statistical modeling of the full wet rainfall spectrum
- No threshold selection to apply Extreme Value Theory
- Joint Extreme Value Theory for both low and high rainfall

Correspondence to:
P. Naveau,
naveau@lsce.ipsl.fr

Citation:
Naveau, P., R. Huser, P. Ribereau, and A. Hannart (2016), Modeling jointly low, moderate, and heavy rainfall intensities without a threshold selection, *Water Resour. Res.*, *52*, 2753–2769, doi:10.1002/2015WR018552.

Received 7 JAN 2016
Accepted 1 MAR 2016
Accepted article online 4 MAR 2016
Published online 9 APR 2016

Modeling jointly low, moderate, and heavy rainfall intensities without a threshold selection

Philippe Naveau[1], Raphael Huser[2], Pierre Ribereau[3], and Alexis Hannart[4]

[1]Laboratoire des Sciences du Climat et de l'Environnement, LSCE/IPSL, CNRS-CEA-UVSQ, Université Paris-Saclay, Gif-sur-Yvette, France, [2]King Abdullah University of Science and Technology, CEMSE Division, Thuwal, Saudi Arabia, [3]Claude Bernard University Lyon 1, Lyon, France, [4]CNRS, Buenos-Aires, Argentina

Abstract In statistics, extreme events are often defined as excesses above a given large threshold. This definition allows hydrologists and flood planners to apply Extreme-Value Theory (EVT) to their time series of interest. Even in the stationary univariate context, this approach has at least two main drawbacks. First, working with excesses implies that a lot of observations (those below the chosen threshold) are completely disregarded. The range of precipitation is artificially shopped down into two pieces, namely large intensities and the rest, which necessarily imposes different statistical models for each piece. Second, this strategy raises a nontrivial and very practical difficultly: how to choose the optimal threshold which correctly discriminates between low and heavy rainfall intensities. To address these issues, we propose a statistical model in which EVT results apply not only to heavy, but also to low precipitation amounts (zeros excluded). Our model is in compliance with EVT on both ends of the spectrum and allows a smooth transition between the two tails, while keeping a low number of parameters. In terms of inference, we have implemented and tested two classical methods of estimation: likelihood maximization and probability weighed moments. Last but not least, there is no need to choose a threshold to define low and high excesses. The performance and flexibility of this approach are illustrated on simulated and hourly precipitation recorded in Lyon, France.

1. Introduction

There exists a wide range of distribution families to statistically model rainfall intensities. For example, *Katz* [1977], *Vrac et al.* [2007], and *Wilks* [2006] argued that most of the precipitation variability can be approximated by gamma distributions. It is, however, also well known [see e.g., *Katz et al.*, 2002] that the tail of the gamma distribution can be too light to capture heavy rainfall intensities. This leads to underestimating return levels and other quantities linked to high quantiles of precipitation amounts. To solve this issue, a popular approach in hydrology [e.g., *Katz et al.*, 2002] is to disregard small and moderate precipitation values and to focus only on the largest rainfall amounts. The advantage of this strategy is that an elegant mathematical framework called *Extreme-Value Theory* (EVT), originating from the pioneering work of *Fisher and Tippett* [1928] and regularly adapted during the last decades [e.g., *de Haan and Ferreira*, 2006], dictates the distribution of heavy precipitation. Specifically, EVT states that rainfall excesses, i.e., amounts of rain greater than a given threshold u, may be approximated by a Generalized Pareto (GP) distribution, provided the threshold and the number of observations are large enough and some mild conditions are satisfied (see section 2 for the GP definition).

Numerous studies [see e.g., *Katz et al.*, 2002; *Cooley et al.*, 2007] have illustrated how the GP distribution can be applied to climate and hydrology sciences. Finding a simple, fast and efficient threshold selection scheme that can provide an optimal threshold remains an elusive task in the realm of hydrological applications (for details, see *Dupuis* [1999] and *Deidda* [2010]). Another obvious drawback is that the GP only models data exceeding a given high threshold, and one can wonder how to model the remaining observations (i.e., lower than the threshold) or equivalently how to deal with the entire range of the data. Recently, there have been a few attempts at modeling the full range of the observations. Carreau and her coauthors [*Carreau and Bengio*, 2009; *Carreau et al.*, 2009; *Carreau and Vrac*, 2011] investigated a semiparametric mixture model that combines hybrid densities built by stitching a Gaussian density with a heavy-tailed GP density. The estimation was performed with a neural network approach and applied in a regression context. This so-called hybrid Pareto model has many advantages, but also two drawbacks. First, it can produce negative values because the low part of the distribution is based on a Gaussian variable, an unwelcome feature for rainfall data. Second, the stitching between the Gaussian and the

Pareto densities is obtained by imposing a strong constraint on the GP and Gaussian parameters. This automatically links the GP shape parameter with the bulk of the distribution, see equations (7) and (8) on page 58 of *Carreau and Bengio* [2009]. In the i.i.d. case, *Frigessi et al*. [2002] proposed another approach based on a mixture model of two components. The first one represents the bulk of the distribution and the second one focuses on the upper tail, with a weight function smoothly connecting the two parts. The Frigessi model may be defined as

$$c[(1-p_{\mu,\tau}(x))\, g_\gamma(x)+p_{\mu,\tau}(x)\, h_\xi(x/\sigma)/\sigma],$$

where $x>0$, c is a normalizing constant, g_γ corresponds to a light-tailed density with parameters γ, and the function h_ξ represents a heavy-tailed GP density with shape parameter $\xi>0$. One of the most interesting aspects of this density mixture is the weight function $p_{\mu,\tau}(\cdot)$ defined by

$$p_{\mu,\tau}(x)=\frac{1}{2}+\frac{1}{\pi}\arctan\left(\frac{x-\mu}{\tau}\right).$$

Since this weight function is nondecreasing such that it takes values in (0, 1] and tends to unity as $x\to\infty$, *Frigessi et al*. [2002] argued that it can play the role of an unsupervised threshold selection algorithm. While *Frigessi et al*. [2002] chose to parametrize the light density g_γ as a Weibull density in their fire loss application, g_γ was a gamma density in the precipitation data studied by *Vrac and Naveau* [2007]. Overall, Frigessi's model has conceptually a lot of advantages; in particular, it removes the delicate choice of a predetermined threshold. In practice, there are important drawbacks. Frigessi's model has a lot of parameters (six) in the simple i.i.d. context and inference is not straightforward. In their simulation study under the true model, *Frigessi et al*. [2002] wrote on page 227 that "for all parameters, the estimates are rather spread. Especially, τ and the Weibull parameters are difficult to estimate, and the estimators are clearly dependent." In addition, their Tables 2 and 3 clearly showed that the GP shape parameter ξ was strongly underestimated (the true value of ξ was above the 75% quantile estimate). Concerning the weight function $p_{\mu,\tau}(x)$, *Vrac and Naveau* [2007] observed those difficulties with rainfall data: the estimates of τ were poor and very close to zero, meaning that the weight function for their rainfall data converged to a step function with a jump at μ. Hence, a discontinuity around the value μ was reintroduced, which is an undesirable feature. Another drawback resides in the constraint of having a strictly positive GP shape parameter, owing to identifiability problems when $\xi=0$. In hydrology, the GP shape parameter can tend to zero when the time scale increases, e.g., say from hourly to weekly rainfall amounts.

The research developed below attempts to address the same issues treated by authors like *Frigessi et al*. [2002], *Carreau et al*. [2009], *Tancredi et al*. [2006], and *Li et al*. [2012], but we would like to avoid the use of mixtures that can quickly inflate the number of parameters (for more details on nonparametric approaches, see *MacDonald et al*. [2011]). A popular road in statistics to increase the flexibility of a given density, here the GP, is to simply multiply it by a simple nonnegative function and renormalize the product to make it a valid pdf. This simple idea is the cornerstone of the so-called *skewed distributions* research field [e.g., *Genton*, 2004]. The main difficulty resides in choosing a multiplicative function that has be to simple enough to keep computational issues at bay, and complex enough to bring a real added value in terms of flexibility. For example, one can carefully choose an appropriate multiplicative function by taking advantage of second-order rates of convergence for a variety of tails [see *Falk et al*., 2010] (see section 2). It is also important to emphasize that this skew-based approach is not the unique way to construct GP distribution extensions.

Recently, *Papastathopoulos and Tawn* [2013] proposed a very interesting and general alternative to build a variety of GP distribution extensions. As any continuous random variable can be generated by applying its inverse cdf to uniform draws (see also the CDF-t transform used in hydrology [e.g., *Kallache et al*., 2011]), one can generate GP-like random variables by replacing the uniform draws by something richer like Beta distributed draws. Again, the main difficulty is to find the right balance between computational simplicity, added flexibility, and desirable features, such as retaining the upper tail behavior. *Papastathopoulos and Tawn* [2013] proposed mainly three types of extensions. Besides establishing the link with skewed distributions, one major difference of the present paper with respect to *Papastathopoulos and Tawn* [2013] is that we take advantage of EVT to model also low rainfall intensities. Although low precipitation amounts are bounded by zero, the lower tail should, in principle, also comply with EVT.

To finish this brief overview about extended GP distributions, we would also like to mention the work of *Beirlant et al*. [2009] who introduced and studied another type of extended GP distribution. As in *Papastathopoulos and Tawn* [2013], their ultimate goal was to improve the estimation of the upper tail shape parameter, and they carefully studied how to select a suitable threshold. Our aim is different in the

sense that we want to model adequately the full range of rainfall, and not just to improve inference for high quantiles. For example, most crop computer models require simulating low, moderate and large precipitation to explore their sensitivity. In summary, our main goal here is to offer a practical model and fast estimation procedures to describe the full precipitation range while bypassing a threshold selection, and not only to improve, per say, the estimation of high quantiles like in *Papastathopoulos and Tawn* [2013] and *Beirlant et al*. [2009]. In particular, low rainfall will also be modeled using the EVT paradigm.

The paper is organized as follows. After recalling a few basic concepts used in EVT, section 2 presents several types of extended GP models and describes a simple sampling scheme. Section 3 discusses inference procedures based on probability weighted moments and maximum likelihood, the performance of which is assessed by an extensive simulation study in section 3.2. Section 4 discusses an application to hourly precipitation in Lyon, France. Finally, we summarize our results and discuss some future research directions in section 5. Before closing this section, we would like to emphasize that we do not treat dry events, but we only focus on positive rainfall intensities.

2. A Rainfall Intensity Model

2.1. Heavy Rainfall Modeling

According to basic univariate EVT [e.g., *Coles*, 2001; *Embrechts et al*., 1997], the probability that large rainfall amounts exceeding a well-chosen high threshold u are larger than x can be approximated by a Generalized Pareto (GP) tail defined as

$$\bar{H}_\xi\left(\frac{x-u}{\sigma}\right),$$

where the survival function $\bar{H}_\xi$ corresponds to

$$\bar{H}_\xi(x)=\begin{cases}(1+\xi x)_+^{-1/\xi}, & \text{if } \xi \neq 0,\\ \exp(-x), & \text{if } \xi=0,\end{cases} \tag{1}$$

with $a_+=\max(a,0)$. The scalar $\sigma>0$ represents the scale parameter. The shape parameter ξ describes the GP tail behavior. If ξ is negative, the upper tail is bounded. If ξ is zero, this corresponds to the case of an exponential distribution, where all moments are finite. If ξ is positive, the upper tail is unbounded but higher moments eventually become infinite. These three cases are termed "short-tailed," "light-tailed," and "heavy-tailed," respectively. The flexibility of the GP distribution to describe three different types of tail behavior makes it a universal tool for modeling excesses. In our case, we assume that heavy rainfall data have either exponential tails ($\xi=0$) or heavy tails ($\xi>0$). This condition appears to be satisfied for most heavy rainfall data [e.g., *Katz et al*., 2002].

Although rarely used in hydrology and climatology, a few approaches of GP distribution extensions have been studied in theoretical statistics. For example, the third edition of the book of *Falk et al*. [2010] describes different extensions of the GP density. For nonnegative GP shape parameters, these authors studied the theoretical properties of densities of the form

$$\text{cst}\times\frac{1}{\sigma}h_\xi\{(x-u)/\sigma\}\left[1+\mathcal{O}\left\{\bar{H}_\xi^\delta(x/\sigma)\right\}\right], \tag{2}$$

where $\delta>0$ and the notation $\mathcal{O}(v)$ means that the ratio $\mathcal{O}(v)/v$ is bounded as v converges to zero; see *Falk et al*. [2010, Proposition 2.2.1]. This class encompasses a broad family of densities. The main idea of equation (2), basically multiplying a density like $h_\xi(.)$ by another function, has been extensively studied in the so-called *skewed distributions* research field [e.g., *Genton*, 2004]. The archetypal example is the skew normal pdf [e.g., *Azzalini*, 1985] defined by the product

$$2\phi(x)\,\Phi(\lambda x)$$

of $\phi(x)$, the standard Gaussian pdf, with its cdf $\Phi(x)$. The parameter λ regulates the skewness, with $\lambda=0$ yielding the normal pdf.

2.2. Low Rainfall Modeling

Before explaining our approach to model the full precipitation range (zeros excluded), we need to address the often overlooked question of how to model low rainfall intensities. At first sight, it may be confusing to try to

apply EVT to low rainfall because they appear to be nonextreme in terms of intensity; they are even bounded below by zero. The beauty of EVT resides in its capacity to also model the conditional random variable $[X-v|X<v]$ given the event $[X<v]$, even when the distribution of $[X-v|X<v]$ does not produce strong intensities itself, i.e., in the case of the GP of Weibull type that applies for low rainfall data. This is possible whenever the threshold v becomes small enough. Mathematically, this can be seen by flipping the sign of the rainfall amounts X, i.e., defining the variable $Y=-X$. The largest negative rainfall in a sample of observations Y may be fitted using a GP distribution with a negative shape parameter, say $-1/\kappa$ with $\kappa \geq 0$, and a positive scale parameter, say $v>0$. Specifically, this means that the upper tail of Y should tend to a GP distribution, i.e.,

$$\mathbb{P}(Y>-x|Y>-v) \approx \bar{H}_{-1/\kappa}\left(\frac{-x+v}{v}\right),$$

as v approaches zero for any x such that $0<x<v$. Obviously, the upper limit of Y is zero, which explains why the shape parameter has to be negative in order to have a short-tailed GP. Furthermore, it also implies that the threshold v has to be chosen such that $\bar{H}_{-1/\kappa}(0)=0$, leading to the constraint $v=\kappa v$. Consequently,

$$\mathbb{P}(Y>-x|Y>-v) \approx \text{cst}\times x^{\kappa}.$$

In other words, this suggests that low rainfall intensities might be adequately described by a power law

$$\mathbb{P}(X \leq x) \approx \text{cst}\times x^{\kappa}, \text{ for any small } x \geq 0.$$

Notice that this condition is satisfied by a gamma density $f(x) \propto x^{\kappa-1}e^{-x/\theta}$, $x \geq 0$, $\kappa, \theta > 0$.

Having different shape parameters for low and heavy rainfall can be justified by the fact that the physical mechanisms associated with high and low rainfall rates are fundamentally different and, therefore, one should not expect both ends of the distribution to exhibit similar properties, i.e., having identical shape parameters.

2.3. Full Range Modeling

According to the two previous sections, we wish that both sides of the rainfall spectrum are in compliance with EVT. Mathematically, this desideratum translates into the following tail approximations

$$\mathbb{P}(X \leq x) \approx \begin{cases} 1-\text{cst}\times\bar{H}_{\xi}\left(\frac{x}{\sigma}\right), & \text{for any ``large'' } x, \\ \text{cst}\times x^{\kappa}, & \text{for any ``small'' } x \text{ near } 0. \end{cases} \quad (3)$$

The widely used gamma density is in agreement with (3) for low values, but fails at representing correctly high values, while the contrary is true for the GP distribution. The exponentially decaying tail of the gamma density typically leads to a drastic underestimation of probabilities of extreme events. We therefore aim at creating a gamma-like density that ressembles a GP density on both tails, while bypassing the threshold selection problem that brings two unwelcome discontinuities between low and moderate, and moderate and heavy rainfall. One strategy could be to create a mixture based on $\bar{H}_{\xi}(x)$ and x^{κ}. This approach is certainly valuable but we do not pursue it here for the following reasons. First, the function x^{κ} needs to be defined on the compact support [0, 1] to be a valid cdf, and the end point of this interval would create a discontinuity in the mixture. Second, the inference might be complex (using an EM algorithm). Third, the computation of return levels (high quantiles) is not explicit. Alternatively, we want to propose a single pdf with the appropriate lower and upper tails and no hidden states. To achieve this goal, we shall follow the footsteps of *Papastathopoulos and Tawn* [2013]. Although these authors focused on the upper tail only, their approach can be somehow adapted for modeling the full rainfall range.

Our main starting point is the classical scheme used to simulate GP distributed random draws via the formula [see e.g., *Robert and Casella*, 2004]

$$\sigma H_{\xi}^{-1}(U), \quad (4)$$

where U represents a random variable uniformly distributed on [0, 1] and H_{ξ}^{-1} corresponds to the inverse GP cdf. A simple way to add flexibility to this simulation scheme is to replace the uniform random draw U in (4) by $V=G^{-1}(U)$, where G is a continuous cdf on [0, 1]. A richer family of random variables may therefore be spanned by defining

$$X=\sigma H_{\xi}^{-1}\{G^{-1}(U)\}. \tag{5}$$

The main question is to find a class of distributions G such that the upper tail behavior with shape parameter ξ is preserved and the cdf of X for values near zero behaves like x^{κ}. If we denote the tail of G by $\bar{G}=1-G$, these constraints can be fulfiled whenever

A.

$$\lim_{v\downarrow 0}\frac{\bar{G}(1-v)}{v}=a, \text{ for some finite } a>0,$$

B.

$$\lim_{v\downarrow 0}\frac{G\{v\,w(v)\}}{G(v)}=b, \text{ for some finite } b>0,$$

where $w(v)$ is any positive function such that $w(v)=1+o(v)$ as $v\to 0$,

C.

$$\lim_{v\downarrow 0}\frac{G(v)}{v^{\kappa}}=c, \text{ for some finite } c>0.$$

To understand these three conditions, one has to notice that equation (5) may be expressed in terms of the cdf $F(x)=\mathbb{P}(X\le x)$ or the tail $\bar{F}(x)=\mathbb{P}(X>x)$ as

$$F(x)=G\left\{H_{\xi}\left(\frac{X}{\sigma}\right)\right\} \text{ and } \bar{F}(x)=\bar{G}\left\{H_{\xi}\left(\frac{X}{\sigma}\right)\right\}. \tag{6}$$

Constraint (A) implies that the upper tail of X is equivalent to a Pareto tail in the sense that the ratio $\bar{F}(x)/\bar{H}_{\xi}(x/\sigma)=\bar{G}(1-v)/v$ with $v=\bar{H}_{\xi}(x/\sigma)$ converges to a constant, as $x\to\infty$. Similarly, the ratio $\frac{F(x)}{G(x)}$ can written as

$$\frac{G\left(\{x/\sigma\}\,\{H_{\xi}(\frac{x}{\sigma})/(x/\sigma)\}\right)}{G(x/\sigma)}\frac{G(x/\sigma)}{G(x)}=\frac{G\{v\,w(v)\}}{G(v)}\frac{G(v)}{G(\sigma v)}$$

where $w(v)=H_{\xi}(v)/v$ with $v=x/\sigma$. The term $\frac{G\{v\,w(v)\}}{G(v)}$ converges to a nonnull constant, as $v\to 0$, because of the constraint (B) and the Taylor expansion $\{1-(1+\xi v)^{-1/\xi}\}/v=1+o(v)$. The ratio $\frac{G(v)}{G(\sigma v)}$ tends to the constant $\sigma^{-\kappa}$ because of (C). In other terms, the constraints (B) & (C) ensure that low values are driven by G. In addition, constraint (C) forces G to behave as a GP of Weibull type for this lower tail.

Before presenting parametric examples of the function G, we would like to emphasize two important practical advantages of (5). This equation provides a straightforward and fast way of simulate random draws from the cdf F whenever the G^{-1} is available. The same is true when return levels have to be computed using the explicit formula

$$x_p=F^{-1}(p)=\begin{cases}\dfrac{\sigma}{\xi}\left[\{1-G^{-1}(p)\}^{-\xi}-1\right], & \text{if } \xi>0,\\ -\dfrac{\sigma}{\xi}\log\{1-G^{-1}(p)\}, & \text{if } \xi=0,\end{cases} \tag{7}$$

$0<p<1$. Practical reasons should drive our choice concerning particular parametric forms of the cdf G.

2.4. Parametric Families

We present now four possible parametric families for $G(v)$, $v\in[0,1]$, which satisfy the above conditions (A), (B) and (C), namely

i. $G(v)=v^{\kappa}$, $\kappa>0$;
ii. $G(v)=pv^{\kappa_1}+(1-p)v^{\kappa_2}$, $\kappa_1,\kappa_2>0$ and $p\in[0,1]$. Without loss of generality, we may assume that $\kappa_1\le\kappa_2$;
iii. $G(v)=1-Q_{\delta}\{(1-v)^{\delta}\}$, $\delta>0$, where Q_{δ} is the cdf of a beta random variable with parameters $1/\delta$ and 2. That is, the cdf Q_{δ} and corresponding pdf q_{δ} are

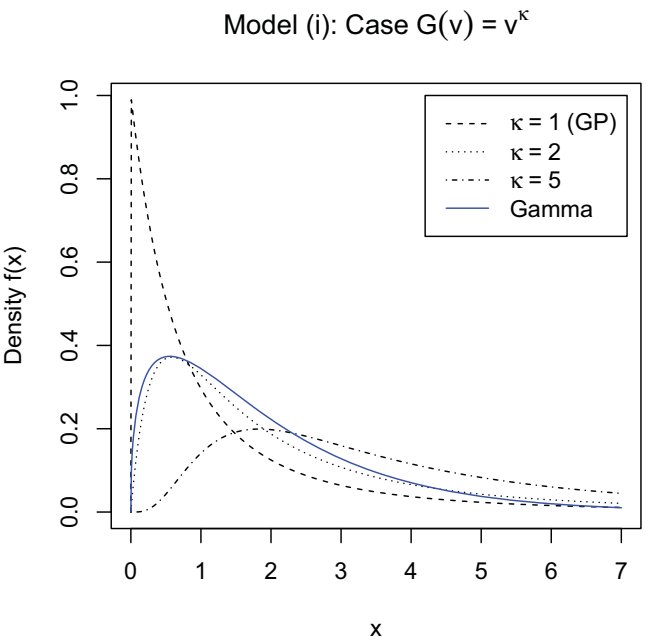

Figure 1. Density function corresponding to Model (6) with $G(v)=v^{\kappa}$, for $\sigma = 1$, $\xi = 0.5$ and lower tail shape parameters $\kappa = 1, 2, 5$ (dashed, dotted, dashed-dotted black curves, respectively). The case $\kappa = 1$ corresponds to the exact GP density. The solid blue curve represents a gamma density with parameters (1.4, 1.4).

$$Q_\delta(v)=\frac{1+\delta}{\delta}v^{1/\delta}\left(1-\frac{v}{1+\delta}\right),$$
$$q_\delta(v)=\frac{1+\delta}{\delta^2}v^{1/\delta-1}(1-v); \quad (8)$$

iv. $G(v)=[1-Q_\delta\{(1-v)^\delta\}]^{\kappa/2}$, $\kappa, \delta > 0$, with Q_δ defined in (8).

Model (i), with the power law distribution $G(v) = v^\kappa$, is clearly the simplest choice and leads to a model in (6) with only three parameters: κ controls the shape of the lower tail, σ is a scale parameter, and ξ controls the rate of upper tail decay. Interestingly, this family corresponds to the type III introduced by *Papastathopoulos and Tawn* [2013]. Figure 1 displays the corresponding density for different lower tail behaviors ($\kappa = 1$, 2, 5) and fixed $\sigma = 1$ and $\xi = 0.5$, compared with a gamma density. The GP model (putting mass $1/\sigma$ at zero) is recovered when $\kappa = 1$, and more flexibility for low values is achieved by varying this parameter. As expected, the gamma and extended GP densities behave similarly for small and moderate values, but the discrepancy increases further in the tail.

A way to increase the flexibility of this fairly simple model is to consider Model (ii), which is a mixture of power laws with $G(v)=pv^{\kappa_1}+(1-p)v^{\kappa_2}$. This model again satisfies the three conditions (A), (B) and (C), with the latter holding true by setting $\kappa = \kappa_1$ (using the convention that $\kappa_1 \le \kappa_2$). This means that the lower tail behavior is controlled by κ_1, whereas κ_2 modifies the shape of the density in its central part. With this specification, Model (6) has five parameters (p, κ_1, κ_2, σ, ξ). Figure 2 illustrates the flexibility of this model with $p = 0.5$, $\kappa_1 = 2$ and different values of κ_2, by comparison with a gamma density.

Model (iii), with $G(v)=1-Q_\delta\{(1-v)^\delta\}$, describes another nontrivial and interesting choice for $G(v)$, which makes a link with the work of *Falk et al.* [2010]. This fairly complex choice of $G(v)$ corresponds in fact to the very simple case where $\mathcal{O}(v)=-v$ and $u = 0$ in (2), i.e.,

$$f(x;\xi,\sigma,\delta)=\frac{1+\delta}{\delta}\,\frac{1}{\sigma}h_\xi(x/\sigma)\left\{1-\bar{H}_\xi^\delta(x/\sigma)\right\}, \quad (9)$$

which is illustrated in Figure 3. It is possible to check that the constraints (A)–(C) are satisfied with $\kappa = 2$. As δ increases to infinity, $f(x; \xi, \sigma, \delta)$ becomes closer to the GP density. Moreover, as explained in section 2.3, the tail behaviors of $f(x; \xi, \sigma, \delta)$ and the GP density $\sigma^{-1}h_\xi(x/\sigma)$ are equivalent for large x: very heavy rainfall is captured in the same fashion for both densities, i.e., through the shape parameter ξ. This can also be justified for this model by noticing that

$$\bar{F}(x;\xi,\sigma,\delta)=\frac{1+\delta}{\delta}\bar{H}_\xi(x/\sigma)\left[1-\frac{1}{1+\delta}\bar{H}_\xi^\delta(x/\sigma)\right]. \quad (10)$$

Consequently, since $\bar{H}_\xi(x)$ converges to zero as x tends to infinity, the upper tail behavior of X is asymptotically equivalent to that of a GP distribution, i.e., for large x,

$$\bar{F}(x;\xi,\sigma,\delta) \sim \frac{1+\delta}{\delta}\bar{H}_\xi(x/\sigma), \;=\; \bar{H}_\xi\left(\frac{x-u_\delta}{\sigma_\delta}\right),$$

where $\sigma_\delta=\sigma c_\delta^\xi$ and $u_\delta=\sigma(c_\delta^\xi-1)/\xi$. In other words, the extra term $\{1-\bar{H}_\xi^\delta(x/\sigma)\}$ in (9) does not affect the extremal index ξ representing the main driver of very extreme events. The parameter δ rather increases

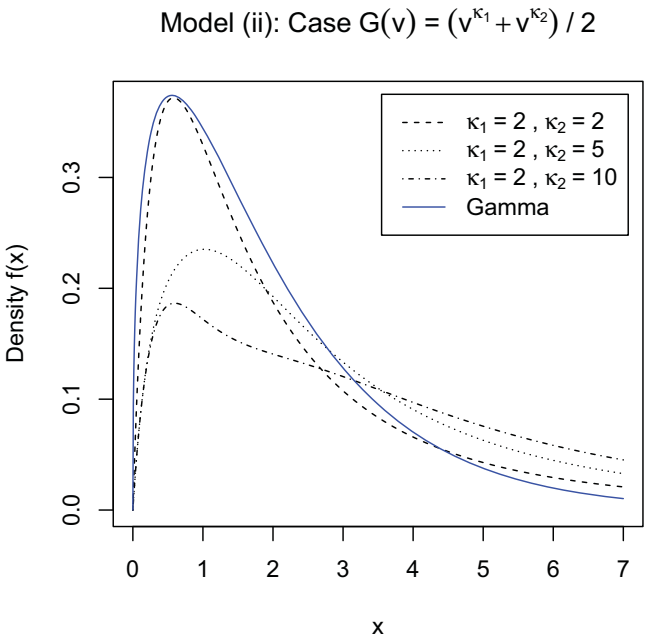

Figure 2. Density function corresponding to Model (6) combined with $G(v)=(v^{\kappa_1}+v^{\kappa_2})/2$ (a special case of Model (ii) with $p = 0.5$), for $\sigma = 1$, $\xi = 0.5$ and parameters $\kappa_1 = 2$, and $\kappa_2 = 2, 5, 10$ (dashed-dotted, dotted, dashed black curves, respectively). The solid blue curve represents a gamma density with parameters (1.4, 1.4).

modeling flexibility for the central part of the distribution. This parameter could be interpreted as a "threshold tuning parameter" that has to be estimated from the data at hand.

Regarding the behavior of $f(x; \xi, \sigma, \delta)$ near zero, we can immediately notice from (9) that $f(0; \xi, \sigma, \delta) = 0$. A drawback of (9) resides in the fact that the Taylor expansion $(1+x)^a \sim 1+ax$ (for small x) implies that the lower tail behavior of $f(x; \xi, \sigma, \delta)$ is

$$f(x;\xi,\sigma,\delta) \sim \frac{1+\delta}{\sigma^2}x, \text{ for } x \text{ near zero.}$$

This means that the lower tail $F(x; \xi, \sigma, \delta)$ is of type x^2 (i.e., $\kappa = 2$ in condition (C)) and consequently, the lower tail behavior is not estimated from the data but imposed by the choice of Model (9).

Model (iv), with $G(v)=[1-Q_\delta\{(1-v)^\delta\}]^{\kappa/2}$, remedies this problem by adding an extra parameter, $\kappa > 0$, controlling the lower tail behavior. This leads to the cdf

$$F(x)=\left[1-Q_\delta\left\{\bar{H}_\xi^\delta\left(\frac{x}{\sigma}\right)\right\}\right]^{\kappa/2}, \tag{11}$$

where κ, δ, ξ describe the low, moderate and upper parts of the distribution, respectively, and σ is a scale parameter. In particular, the lower and upper tails are, by construction, GP with shape parameters κ and ξ, respectively.

The inverse function $G^{-1}(u)$ is available in closed-form for Model (i), may be easily computed using numerical inversion for Model (ii), and is based on the quantile function $Q_\delta^{-1}(u)$ of the Beta distribution with parameters $1/\delta$ and 2 for Models (iii) and (iv). When combined with the formula (7), this makes the simulation and the computation of return levels straightforward and fast, which is important from a hydrological point of view. To simulate from Model (6), one simply needs to randomly draw a uniform variable U in [0, 1], and then to apply the quantile function (7) as $X=F^{-1}(U)$.

It is worth mentioning that all models described in this section are eligible as valid parametric families fulfiling (3), though Model (iii) has restricted flexibility in its lower tail. The choice of the most appropriate model depends on the application at hand and classical devices like quantile-quantile plots, histograms, AIC and so on, can guide the hydrologist to opt for a specific model.

3. Inference

3.1. Estimation of Unknown Parameters

A variety of inference methods exists and, for heavy rainfall analysis, two options are popular among hydrologists: maximum likelihood (ML) estimation and a method of moments based on probability weighted moments (PWMs). In this paper, we investigate how these two classical inference techniques can be implemented within our framework.

Concerning the ML approach, the likelihood function may be easily obtained from (6) whenever the function $G(v)$ is easily differentiable. This is the case for the parametric families introduced in section 2.4, and more details are given in Appendix A.

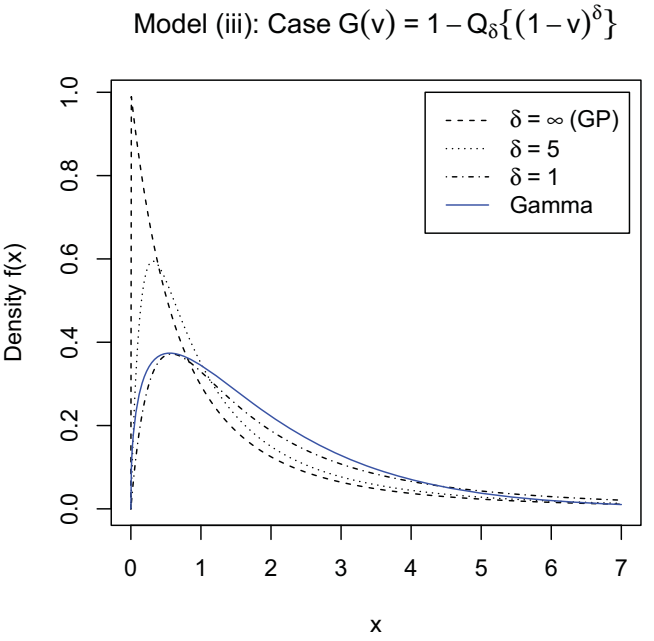

Figure 3. Density function corresponding to Model (6) combined with $G(v)=1-Q_\delta\{(1-v)^\delta\}$, for $\sigma=1$, $\xi=0.5$ and parameters $\delta=1, 5, \infty$ (dashed-dotted, dotted, dashed black curves, respectively). The limiting case $\delta=\infty$ corresponds to the exact GP density. The solid blue curve represents a gamma density with parameters (1.4, 1.4).

As for the PWMs approach, it has a long tradition in statistical hydrology [e.g., *Landwehr et al.*, 1979; *Hosking and Wallis*, 1987]. It has been recently revisited by statisticians [e.g., *Ferreira and de Haan*, 2014] and applied in various settings [e.g., *Naveau et al.*, 2014]. A recent study of *Caeiro and Gomes* [2011] theoretically compares different estimation methods for the shape parameter ξ. Besides its simplicity, the PWMs approach usually performs reasonably well compared to other estimation procedures. Additional arguments in favor of PWMs are that they are typically quickly computed, even in nonstationary contexts [e.g., *Naveau et al.*, 2014], and relatively robust against model misspecification.

The idea of the PWMs approach is simple: for a pdf specified by s parameters, we need to find the explicit expressions of s (weighted) moments that are function of these parameters. Having s empirical moments and s theoretical moments (depending on s unknown parameters), we can pursue a method-of-moments by equating these quantities and solving the resulting system of equations [e.g., *Diebolt et al.*, 2008, 2007].

For the model proposed in (6), it is convenient to work with PWMs of the form

$$\mu_s=\mathbb{E}\left(X\bar{F}^s(X)\right), \quad s=0,1,\ldots. \tag{12}$$

Appendix B provides the explicit PWMs for the parametric families (i), (ii) and (iii) defined by $G(v)=v^\kappa$, $G(v)=pv^{\kappa_1}+(1-p)v^{\kappa_2}$ and $G(v)=1-Q_\delta\{(1-v)^\delta\}$, respectively. Although PWMs are available in closed form for these cases, estimated parameters cannot be expressed as functions of PWMs in general. In practice, this limitation does not cause any particular problem because statistical softwares like the R package gmm [*Chaussé*, 2010] provides numerical solutions to such method-of-moments nonlinear systems of equations. Confidence intervals can also be obtained from this R package. The extension to the case $G(v)=[1-Q_\delta\{(1-v)^\delta\}]^{\kappa/2}$ with Model (iv) is, however, more complicated and PWMs have to be computed by Monte Carlo simulations using a large number of replicates. For Model (ii), the estimation of the five parameters σ, ξ, p, κ_1 and κ_2 requires PWMs of orders $s=0$, 1, 2, 3, 4, while for Model (iv), only the first four moments can be used to estimate σ, ξ, κ and δ. For Models (i) and (iii), only the first three moments are needed.

The following section illustrates how the inference performs.

3.2. Simulation Study

In this section, we assess the performance of the PWMs and ML estimators by simulation. All results presented here are based on the following setting. The scale parameter in Model (6) is always set to one ($\sigma=1$). The shape upper tail parameters can take three values, $\xi=0.1$, 0.2, 0.3, classical values for precipitation data. The sample size is fixed to $n=300$, and 10^5 replicates are used to compute basic statistical metrics like root mean squared errors (RMSEs). Different settings provide similar conclusions. We explore the four proposed models, precisely using the following parameter values

1. $G(v)=v^\kappa$, with lower tail parameter $\kappa\in\{1, 2, 5, 10\}$;
2. $G(v)=pv^{\kappa_1}+(1-p)v^{\kappa_2}$, with $p=0.4$, $\kappa_1\in\{1, 2, 5, 10\}$ and $\kappa_2\in\{2, 5, 10, 20\}$,
3. $G(v)=1-Q_\delta\{(1-u)^\delta\}$, with skewness parameter $\delta\in\{0.5, 1,2, 5\}$,

4. $G(v)=[1-Q_\delta\{(1-v)^\delta\}]^{\kappa/2}$, with $\delta \in \{0.5, 1, 2, 5\}$, and $\kappa \in \{1, 2, 5, 10\}$,

   with starting values for the numerical solvers fixed arbitrarily to $\sigma = 2$, $\xi = 0.15$, $\kappa = 3$, $\delta = 1.5$, $p = 0.5$, $\kappa_1 = 3$ and $\kappa_2 = 4$.

To summarize the performance of the ML and PWMs estimators, Figure 4 displays boxplots of estimated parameters and 99%-quantiles (i.e., 1/100 return levels) for representative cases, see the caption for the exact parameters values (Model (iv) gives similar results). Skewness parameters like δ or κ_2 are difficult to estimate by ML, as it was noticed by *Sartori* [2006] for the skew-normal case and *Ribereau et al*. [2015] for the extended GEV case. Figure 4 also shows generally higher variability of PWMs estimators compared to ML estimators, especially for tail decay parameters (ξ or κ) and high quantiles. More importantly, both inferential methods provide reasonable estimates for a moderate sample size of $n = 300$, even with a five-parameter model like Model (ii) (though the ML estimation of $\kappa_2 = 5$ is quite poor).

To fine tune our comparison between these two classical estimation approaches, Table 1 reports the ratio of RMSEs of PWMs and ML estimators for the four models. The ratio of RMSEs for the estimated 99%-quantile is reported in bold. Values lower than unity indicate that PWMs estimators perform better, and vice versa.

Overall, PWMs and ML estimators have a similar performance, though the latter are generally slightly better. This is especially the case for Model (i) with large $\kappa \geq 5$, although high quantiles may still be reasonably well estimated by PWMs (see the case $\kappa = 10$, $\xi = 0.1$). The case of Model (ii) exacerbates that five parameters imply a much higher instability, especially for $\kappa_2 = 2$ by ML and for $\kappa_1 = 10$ by PWMs. Model (iii) emphasizes one more time that PWMs perform well for the skewness parameter δ, but interestingly, high quantiles seem to be better estimated using the ML approach. Similar conclusions apply for 98% and 99.5%-quantiles. For Model (iv), ML estimators slightly but constantly outperform PWMs in terms of RMSE for high quantiles. This is surprising because the skewness parameter δ has much higher variability with the ML approach.

One may wonder if imposing a parametric model on the whole data set "deteriorates" the fit of the largest values that could be obtained by a classical GP approach based on a small fraction of extreme data. To assess this, we simulated data from Model (i) with $\sigma = 1$, $\xi = 0.2$, $\kappa = 2$ (based on the same setting as before), and estimated the upper tail parameter ξ and the 99%-quantile (i.e., the 1/100 return level), either by fitting the true model to the whole data set by ML or by fitting the GP distribution to excesses above the 95%-quantile. In a nutshell, the full range modeling approach improves the estimate of ξ, the ratio of RMSEs being equal to 3.22, while this ratio is only 1.12 for the 99%-quantile, indicating that the gain is weaker for high (though not extremely high) quantiles. This simple experiment suggests that the bulk of the distribution may help to estimate upper (respectively lower) tail parameters or high (respectively low) quantiles, provided that the assumed model is, at best, the correct one or, at least, flexible enough to adapt itself to the data at hand.

4. Hourly Rainfall in Lyon (France)

As an illustrative example, we analyze hourly precipitation from 1996 to 2011 recorded at the French weather station of Lyon. There is a clear seasonal signal in terms of cumulative precipitation in which Fall (September–October–November) is very wet, as well as Spring (March–April–May) but to a lesser degree. Summer (June–July–August) is much drier and Winter (December–January–February) has low cumulative rainfall but can experience strong and short length episodes. For this reason, we treat the seasons as individual data sets.

The four different extended GP models described in section 2 were separately fitted to rainfall intensities for each season using ML and PWMs, as outlined in section 3. To reduce temporal short-term dependence, every third observation was retained for the analysis of each time series. After removing the dry events (i.e., zero values), the sample sizes are equal to 945 (Spring), 715 (Summer), 1073 (Fall) and 925 (Winter); see the histograms in Figure 5. According to the Akaike information criterion (AIC) and simple graphical diagnostics, Model (i) with $G(v)=v^\kappa$ in (6) performs best overall. The fit of Model (ii) with $G(v)=pv^{\kappa_1}+(1-p)v^{\kappa_2}$ is usually quite reasonable, too, but this model adds two parameters (i.e., it has five parameters in total) and the AIC is thus often in favor of Model (i). As for Model (iii) with $G(v)=1-Q_\delta\{(1-v)^\delta\}$, it lacks flexibility in the lower tail and thus constantly appears to be the worst. Model (iv) with $G(v)=[1-Q_\delta\{(1-v)^\delta\}]^{\kappa/2}$ is often

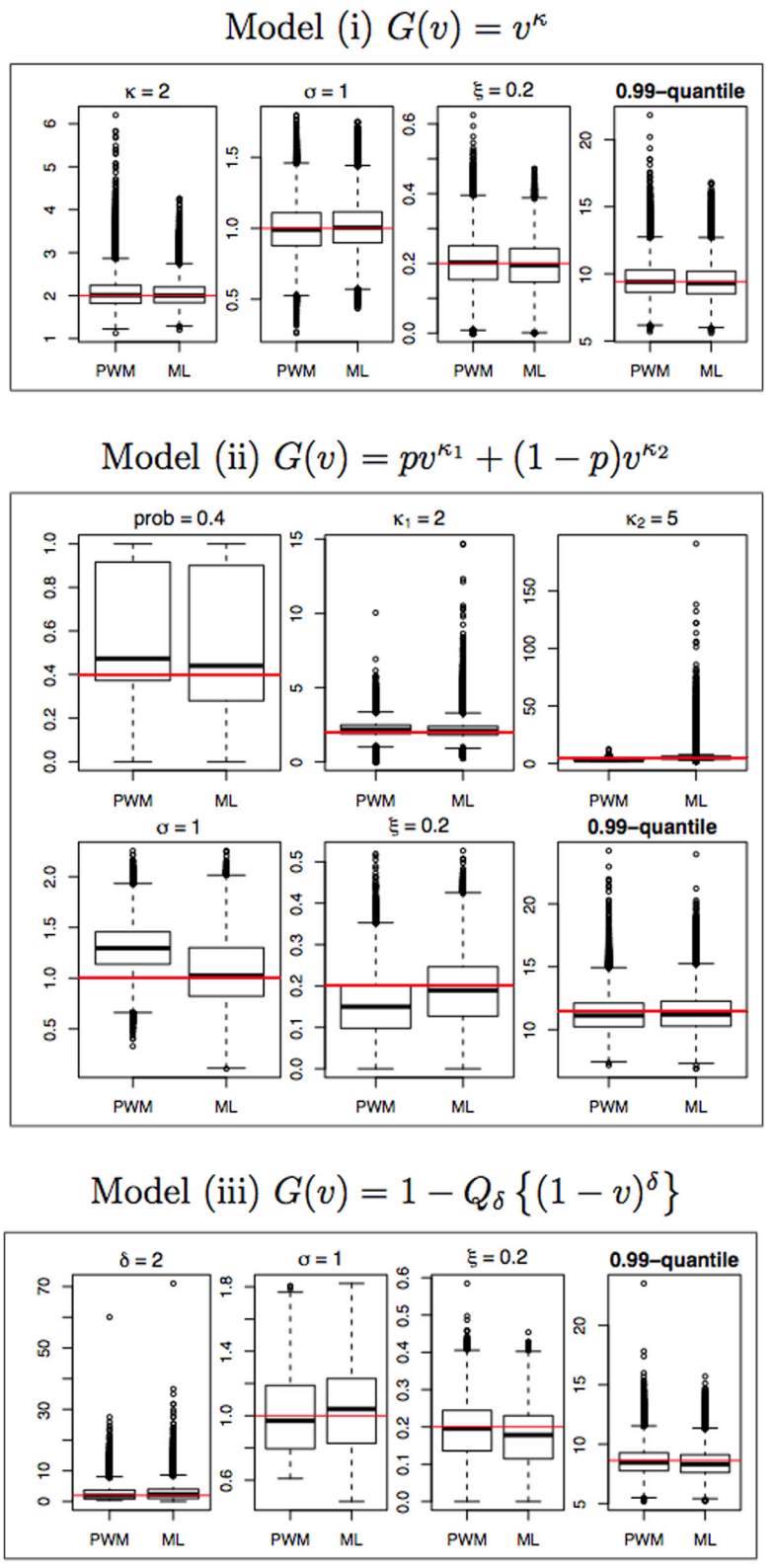

Figure 4. Boxplots of estimated parameters and 0.99-quantiles (i.e., 1/100 return levels) using (left) PWMs and (right) ML, for Model (i) with $\sigma = 1$, $\xi = 0.2$, $\kappa = 2$, model (ii) with $p = 0.4$, $\kappa_1 = 2$, $\kappa_2 = 5$, $\sigma = 1$ and $\xi = 0.2$, and Model (iii) with $\sigma = 1$, $\xi = 0.2$, $\delta = 2$. Boxplots are based on 10^5 independent replicates, and true values are represented by horizontal red lines.

Table 1. Ratio of Root Mean Squared Errors (RMSEs) of PWMs and MLE Based on Estimates Obtained From 10^5 Independent Data Sets of Size $n = 300$[a]

Model (i) $G(v) = v^{\kappa}$			
κ	ξ		
	0.1	0.2	0.3
1	1.06/1.02/1.17	1.06/0.98/1.19	1.11/1.00/1.27
	1.01	**0.98**	**0.98**
2	1.05/1.02/1.13	1.07/1.00/1.18	1.15/1.05/1.33
	1.01	**0.99**	**1.02**
5	1.74/1.39/1.29	1.68/1.49/1.31	1.19/1.09/1.34
	1.00	**1.14**	**1.04**
10	2.40/1.70/1.41	4.21/2.82/1.32	1.39/1.26/0.69
	0.76	**1.26**	**1.06**
Model (iii) $G(v)=1-Q_{\delta}\{(1-v)^{\delta}\}$			
δ	ξ		
	0.1	0.2	0.3
0.5	1.00/1.03/0.98	0.91/0.90/0.91	0.92/0.92/0.91
	1.02	**1.02**	**1.05**
1	0.95/0.99/0.94	0.89/0.90/0.90	0.90/0.92/0.90
	1.03	**1.02**	**1.05**
2	1.02/1.00/0.95	0.95/0.93/0.91	0.93/0.94/0.92
	1.03	**1.01**	**1.04**
5	1.06/1.05/0.95	0.99/0.97/0.98	0.98/0.97/1.07
	1.02	**0.98**	**0.99**
Model (ii) $G(v)=pv^{\kappa_1}+(1-p)v^{\kappa_2}$			
κ_1	κ_2		
	2	5	10	20
1	**0.57**	**0.96**	**0.68**	**0.74**
2		**0.95**	**0.98**	**0.71**
5			**1.29**	**1.04**
10				**1.38**
Model (iv) $G(v)=[1-Q_{\delta}\{(1-v)^{\delta}\}]^{\kappa/2}$				
κ	δ			
	0.5	1	2	5
1	**1.13**	**1.13**	**1.09**	**1.02**
2	**1.09**	**1.09**	**1.06**	**1.01**
5	**1.10**	**1.09**	**1.07**	**1.04**
10	**1.02**	**1.13**	**1.03**	**1.07**

[a]Each cell in bold represents the ratio of RMSEs for the 99%-quantile. Nonbold cells correspond to ratios for each parameter, i.e., $\sigma/\xi/\kappa$ for Model (i), and $\sigma/\xi/\delta$ for Model (iii). Numbers lower than one indicate that the PWMs estimator performs better, and vice versa.

comparable to Model (i), though the estimation of the skewness parameter δ is sometimes quite poor. For these reasons, Figure 5 and Table 2 report the results of Model (i) only.

In the simulation study in section 3.2, the ML approach was found to have generally lower RMSEs than PWMs. However, these results were based on the assumption that the models fitted were well-specified. When dealing with real data, however, this strong assumption is no longer valid, and we have found that PWMs are (much) more robust against model misspecification since they are based on useful summary statistics, rather than the exact values of observations. This is especially important when a model is fitted to the full range of hourly rainfall data, because the discretization due to instrumental precision (see Figure 6) strongly affects low values, and consequently the estimation of parameters if this feature is not properly taken into account. In particular, the upper tail shape parameter ξ is constantly over-estimated using the classical ML approach, and the effect of the discretization increases as the distribution becomes more concentrated around zero. To counteract this undesirable effect, one might consider two possibilities:

a. one may treat the data as being left-censored. For example, by censoring data below the threshold $x_L = 0.5$ mm, the effective sample size fully contributing to the likelihood becomes 448, 383, 585 and 391 for Spring, Summer, Fall and Winter, respectively;

Table 2. Estimated Parameters for Model (i) With $G(v)=v^{\kappa}$ in (5) Fitted to the Hourly Rainfall Data by Maximum Likelihood (ML) and by Probability Weighted Moments (PWMs); Estimators Censoring Observations Below 0.5 mm Have the Suffix "-c"[a]

Spring Data			
Method	Estimated Parameters		
	κ	σ	ξ
ML	$1_{(0.6,2)} \times 10^3$	$0.00_{(0.00,0.00)}$	$0.81_{(0.78,0.85)}$
ML-c	$0.59_{(0.51,0.73)}$	$1.44_{(1.09,1.65)}$	$0.03_{(0.00,0.14)}$
PWMs	$0.61_{(0.56,0.67)}$	$1.47_{(1.27,1.61)}$	$0.03_{(0.00,0.11)}$
PWMs-c	$0.50_{(0.44,0.57)}$	$1.55_{(1.31,1.73)}$	$0.00_{(0.00,0.08)}$
Full Data			
Method	Estimated Parameters		
	κ	σ	ξ
ML	$1_{(0.5,2)} \times 10^3$	$0.00_{(0.00,0.00)}$	$0.85_{(0.81,0.88)}$
ML-c	$0.80_{(0.64,1.02)}$	$1.00_{(0.73,1.34)}$	$0.29_{(0.16,0.40)}$
PWMs	$0.66_{(0.60,0.72)}$	$1.32_{(1.11,1.53)}$	$0.20_{(0.11,0.28)}$
PWMs-c	$0.62_{(0.52,0.75)}$	$1.22_{(0.93,1.48)}$	$0.22_{(0.12,0.31)}$
Summer Data			
Method	Estimated Parameters		
	κ	σ	ξ
ML	$0.7_{(0.4,2)} \times 10^3$	$0.00_{(0.00,0.00)}$	$0.91_{(0.87,0.96)}$
ML-c	$0.51_{(0.41,0.64)}$	$1.94_{(1.36,2.65)}$	$0.18_{(0.03,0.33)}$
PWMs	$0.56_{(0.51,0.63)}$	$1.82_{(1.46,2.16)}$	$0.20_{(0.11,0.30)}$
PWMs-c	$0.43_{(0.37,0.53)}$	$2.16_{(1.61,2.66)}$	$0.12_{(0.01,0.25)}$
Winter Data			
Method	Estimated Parameters		
	κ	σ	ξ
ML	$3_{(2,8)} \times 10^3$	$0.00_{(0.00,0.00)}$	$0.67_{(0.63,0.70)}$
ML-c	$0.84_{(0.64,1.14)}$	$0.63_{(0.45,0.84)}$	$0.23_{(0.09,0.34)}$
PWMs	$0.59_{(0.54,0.65)}$	$1.07_{(0.92,1.17)}$	$0.04_{(0.00,0.18)}$
PWMs-c	$0.63_{(0.46,0.82)}$	$0.73_{(0.52,1.04)}$	$0.20_{(0.00,0.33)}$

[a]95%-confidence intervals (subscripts) are obtained from 500 nonparametric bootstrap replicates.

b. one may assume that data observed in the interval $[x, x+0.1)$ have been rounded to x and modify the likelihood function accordingly.

A combination of (a) and (b) is also possible. Details about the implementation of these censored likelihoods are given in Appendix A. Here, the left-censoring approach in (a) provides the best results. To compare censored and noncensored approaches, we also provide in Appendix B the explicit expressions of censored PWMs of the form

$$\mu_{s;x_L,x_U} = \mathbb{E}\left(X\bar{F}^s(X) | x_L < X < x_U\right), \quad s=0,1,\ldots. \tag{13}$$

for Models (i), (ii) and (iii). Similarly to noncensored PWMs in (12), explicit expressions are not available for Model (iv) and simulations are needed in this case to calculate (13). Of course, the expectation in (13) boils down to (12) as $x_L \to 0$ and $x_U \to \infty$.

Table 2 reports the estimated parameters obtained by ML and PWMs, based on the naive and left-censored approaches (with suffix "-c"), with 95%-confidence intervals constructed from 500 nonparametric bootstrap replicates. One can clearly see the improvement when data are censored below 0.5 mm for ML estimation: censoring significantly improves the inference, and corrects the estimation of the shape parameters and return levels. By contrast with ML, the approach based on PWMs is much less sensitive to departures from the assumed model. Indeed, the difference between censored and noncensored PWMs estimates is mild (except perhaps for the Winter data, where the estimated shape parameter is underestimated for the noncensored approach). This illustrates the robustness of PWMs in misspecified settings and the sensitivity of ML to the discretization of low rainfall values. Furthermore, there is an overall good agreement between censored likelihood and PWMs, though the ML-c approach has a tendency to slightly overestimate the shape parameters (κ and ξ) with respect to PWMs-c.

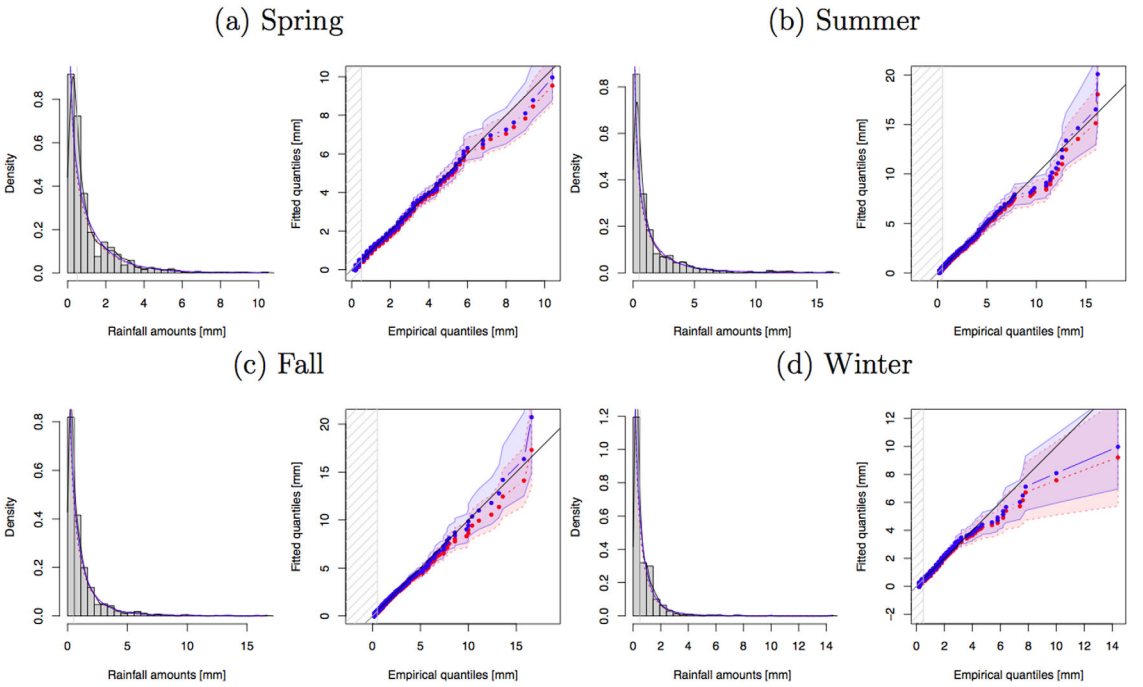

Figure 5. Hourly precipitation (1996–2011, Lyon, France). For each season ((a) Spring, (b) Summer, (c) Fall and (d) Winter), (left) the empirical histogram (grey) for the data. A kernel-based density (black) and fitted extended GP densities (by ML-c in solid blue, by PWMs-c in dashed red) are superimposed; data below 0.5 mm (shaded areas) were censored in the estimation procedure. (right) The corresponding quantile-quantile plots with associated (pointwise) 95% confidence intervals based on 500 bootstrap replicates.

Figure 5 displays the fitted densities for ML-c and PWMs-c approaches, as well as quantile-quantile plots representing the quality of the fit. For the Spring, Summer, and Fall data, the fits appear to be very good, whereas there is some discrepancy at extreme levels for the Winter data. This suggests that there is a need for models that are more flexible in the bulk of the distribution (e.g., mixture models), which could better account for different weather types.

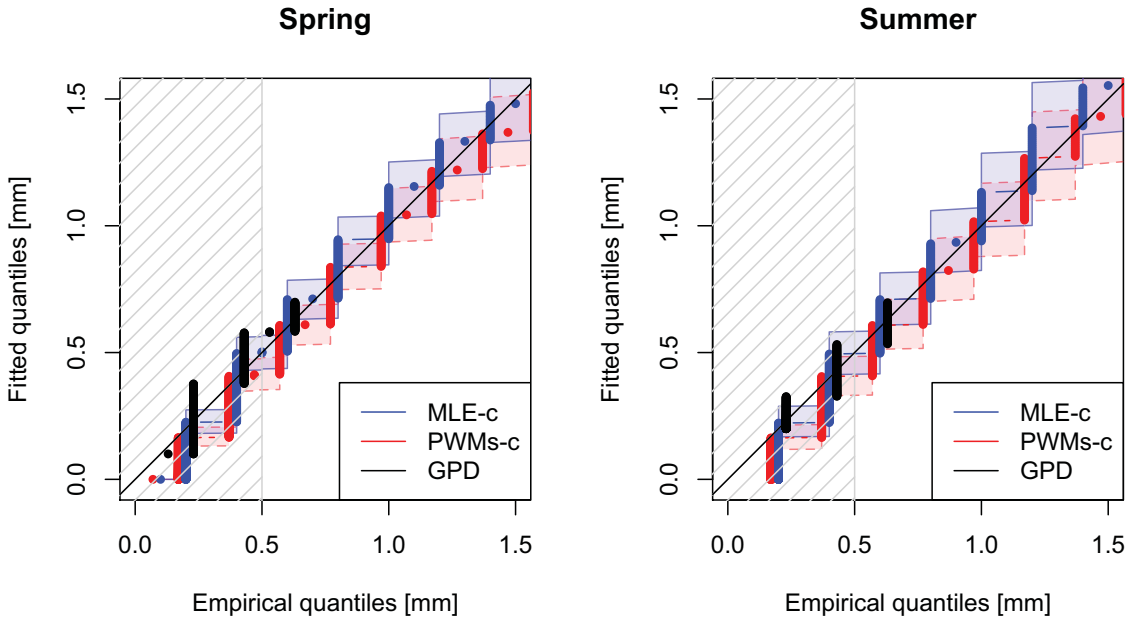

Figure 6. Quantile-quantile plots, zoomed in the lower left region corresponding to low rainfall values (less than 1.5 mm), for the fit of Model (i) (by MLE-c, blue, and by PWMs-c, red, with censoring applied on the shaded area) and the GP distribution (black) to the (left) Spring and (right) Summer rainfall data recorded in Lyon. The fit of Model (i) (with 3 parameters) was performed on the entire data set (censored below 0.5 mm), while the fit of the GP distribution (with three parameters + threshold) used only the rainfall data not exceeding 0.7 mm. The black and red dots have been slightly off-set for better readability.

To visualize the fit on small rainfall, Figure 6 zooms on the Spring and Summer QQ-plots of Figure 5 and compares Model (i) fitted to the entire data set with a classical GP distribution fitted exclusively on low rainfall data not exceeding 0.7 mm. The blue color corresponds to the 95% confidence intervals obtained by ML-c, while the red color represents the confidence intervals by PWMs-c. The three different fits appear to capture well small rainfall extremes, despite the obvious and intrinsic discrete nature of small rainfall observations due to the instrumental rounding at 0.1 mm. The close agreement between the GP model and Model (i) suggests that our approach has the same capacity at representing small rainfall than a GP fit, but with the advantage to also model well moderate and heavy rainfall (see Figure 5).

5. Conclusions and Perspectives

In this work, we attempt to show that it may be possible to model jointly low, moderate and heavy rainfall intensities, without having to fix a threshold. Such a strategy coupled with two classical inference approaches has many practical advantages. It is fast to implement, simple to interpret with only a few parameters and in compliance with EVT for both the upper and lower tails. Concerning the inference, PWMs and ML approaches perform adequately on simulated data, with a slight advantage for ML for high quantiles inference. PWMs seem to be better for skewness. With observed rainfall recordings, the PWMs method appears to be more straightforward and robust, the ML estimator having to be fine-tuned to handle the censoring at 0.5 mm due to precision errors. In any case, both approaches are fast and we advice to use both and compare estimates. It would be of interest to implement a penalized likelihood or a Bayesian approach: putting an informative prior on δ and ξ could certainly improve the analysis for these two parameters that are never simple to infer.

Our object of study was rainfall and consequently, we limit our GP distribution shape parameter to be nonnegative. For other applications, it would be interesting to relax this hypothesis. Regarding precipitation data, obvious extensions could be explored. In this work, we did not address the important issue of modeling rainfall occurrences (i.e., wet or dry events). This leads to the question of how to couple Bernoulli type events with continuous intensities [see e.g., *Koch and Naveau*, 2015]. In the same vein, the modeling of rainfall amounts at multisites remains a statistical challenge and it would be of interest to develop a multivariate version of our proposed distributions.

One drawback could be that we impose a specific form for moderate precipitation. In cases for which this limitation is too stringent, it is always possible to couple our method with a weather-type approach, i.e., by assuming that daily rainfall has to belong to a weather type driven by some specific atmospheric pattern [see e.g., *Ailliot et al.*, 2015]. Hence, our distributions could be fitted for each weather type and consequently, a mixture of pdfs would represent the whole rainfall spectrum. One could even impose the weather-type mixture on the distribution $G(v)$; see Model (iii) in section 3.2.

Appendix A: Likelihood and Censored Likelihood Estimation for Model (6)

From (6), it can be deduced that if $G(v)$ is differentiable and has density $g(v)$, then

$$f(x)=\frac{\partial}{\partial x}F(x)=\frac{\partial}{\partial x}G\left\{H_\xi\left(\frac{x}{\sigma}\right)\right\}=\frac{1}{\sigma}h_\xi\left(\frac{x}{\sigma}\right)g\left\{H_\xi\left(\frac{x}{\sigma}\right)\right\},$$

where h_ξ is the GP density with shape parameter ξ defined in (1). Therefore, for a data set $x_1,\ldots,x_n \overset{iid}{\sim} F$, the likelihood function may be written as

$$L(\psi)=\prod_{i=1}^{n} f(x_i)=\sigma^{-n}\prod_{i=1}^{n} h_\xi\left(\frac{x_i}{\sigma}\right)g\left\{H_\xi\left(\frac{x_i}{\sigma}\right)\right\},$$

where ψ denotes the vector of unknown parameters. The left-censoring approach (a) described and applied in section 4 consists in using instead the censored likelihood

$$L_c^1(\psi)=\prod_{i:x_i<C} F(C)\prod_{i:x_i\geq C} f(x_i)=\prod_{i:x_i<C} G\left\{H_\xi\left(\frac{C}{\sigma}\right)\right\}\prod_{i:x_i\geq C}\sigma^{-1}h_\xi\left(\frac{x_i}{\sigma}\right)g\left\{H_\xi\left(\frac{x_i}{\sigma}\right)\right\},$$

where C is the censoring threshold. If needed, right censoring may be dealt with similarly. The second censoring approach (b) which treats the data as being rounded to the closest 0.1 mm may be based on the following piecewise-censored likelihood function

$$L_c^2(\psi)=\prod_{i=1}^{n}\{F(x_i+0.1)-F(x_i)\}=\prod_{i=1}^{n}\left[G\left\{H_\xi\left(\frac{x_i+0.1}{\sigma}\right)\right\}-G\left\{H_\xi\left(\frac{x_i}{\sigma}\right)\right\}\right].$$

To obtain estimated parameters $\hat{\psi}$, one may maximize $L(\psi)$, $L_c^1(\psi)$ or $L_c^2(\psi)$.

Appendix B: PWMs of the Different Parametric Families Defined in Section 2.4

In this section, we use the notation B(a, b) and IB(x; a, b), $x \in [0, 1]$, to denote the beta and incomplete beta functions with parameters a and b, respectively, i.e.,

$$\mathrm{B}(a,b)=\int_0^1 u^{a-1}(1-u)^{b-1}\mathrm{d}u, \qquad \mathrm{IB}(x;a,b)=\int_0^x u^{a-1}(1-u)^{b-1}\mathrm{d}u.$$

For simplicity, we shall also write for $0 \le x \le y \le 1$, $\mathrm{IB}(x,y;a,b)=\mathrm{IB}(y;a,b)-\mathrm{IB}(x;a,b)=\int_x^y u^{a-1}(1-u)^{b-1}\mathrm{d}u$.

Let $F_L=F(x_L)$, $F_U=F(x_U)$, $\Delta_F=F_U-F_L$, $H_L=H_\xi\left(\frac{x_L}{\sigma}\right)$, and $H_U=H_\xi\left(\frac{x_U}{\sigma}\right)$. From (7), it can easily be deduced that for $s = 0, 1, 2,\ldots$, censored PWMs of the form (13) for Model (6) with $F(x) = G\{H_\xi(x/\sigma)\}$ may be written equivalently as

$$\begin{aligned}\mu_{s;x_L,x_U}&=\mathbb{E}\left(X\bar{F}^s(X)|x_L<X<x_U\right)\\&=\frac{\sigma}{\xi}\left(\mathbb{E}\left[\left\{1-G^{-1}(U)\right\}^{-\xi}(1-U)^s|F_L<U<F_U\right]-\frac{(1-F_L)^{s+1}-(1-F_U)^{s+1}}{(s+1)\Delta_F}\right)\\&=\frac{\sigma}{\xi}\left(\underbrace{\mathbb{E}\left[(1-V)^{-\xi}\{1-G(V)\}^s|H_L<V<H_U\right]}_{E_s}-\frac{(1-F_L)^{s+1}-(1-F_U)^{s+1}}{(s+1)\Delta_F}\right),\end{aligned}$$

where $U \sim \mathrm{Unif}(0, 1)$ and $V \sim G$. Notice that without censoring, the conditional expectation E_s in the above expression reduces to $\mathbb{E}[(1-V)^{-\xi}\{1-G(V)\}^s]$ and $F_L = 0$, $F_U = 1$, $\Delta_F = 1$. Thus, noncensored PWMs μ_s in (12) can easily be obtained from these expressions. In particular, when $G(v) = v$ (i.e., F is the GP distribution), one has

$$E_s=\mathbb{E}[(1-V)^{s-\xi}]=\frac{1}{s-\xi+1} \quad \Rightarrow \quad \mu_s=\frac{\sigma}{(s-\xi+1)(s+1)}, \tag{B1}$$

from which we get $\xi = (\mu_0 - 4\mu_1)/(\mu_0 - 2\mu_1)$ and $\sigma = \mu_0(1 - \xi)$. In this case, PWMs estimates of σ and ξ can be obtained explicitly, but this is no longer possible for the models described below. We next compute the conditional expectation E_s explicitly for Models (i), (ii) and (iii).

For Model (i) with $G(v)=v^\kappa$, one has for any nonnegative integer s

$$\begin{aligned}E_s&=\sum_{j=0}^{s}\binom{s}{j}(-1)^j\mathbb{E}\left\{(1-V)^{-\xi}V^{\kappa j}|H_L<V<H_U\right\}\\&=\frac{\kappa}{\Delta_F}\sum_{j=0}^{s}\binom{s}{j}(-1)^j\,\mathrm{IB}\{H_L,H_U;(j+1)\kappa,1-\xi\}.\end{aligned} \tag{B2}$$

It follows that the first three noncensored PWMs are equal to

$$\mu_0=\frac{\sigma}{\xi}\{\kappa\mathrm{B}(\kappa,1-\xi)-1\}, \qquad \mu_1=\frac{\sigma}{\xi}\left[\kappa\{\mathrm{B}(\kappa,1-\xi)-\mathrm{B}(2\kappa,1-\xi)\}-\frac{1}{2}\right],$$
$$\mu_2=\frac{\sigma}{\xi}\left[\kappa\{\mathrm{B}(\kappa,1-\xi)-2\mathrm{B}(2\kappa,1-\xi)+\mathrm{B}(3\kappa,1-\xi)\}-\frac{1}{3}\right].$$

From (B2), similar expressions may be obtained for censored PWMs.

For Model (ii) with $G(v)=pv^{\kappa_1}+(1-p)v^{\kappa_2}$, one can follow the same lines and obtain that for any nonnegative integer s

$$E_s = \frac{1}{\Delta_F}\sum_{j=0}^{s}\sum_{k=0}^{j}\binom{s}{j}\binom{j}{k}(-1)^j p^k (1-p)^{j-k} A_{j,k},$$

where the coefficients $A_{j,k}$ may be expressed in terms of the incomplete beta function as

$$A_{j,k} = p\kappa_1\,\mathrm{IB}\{H_L, H_U; \kappa_1(k+1)+\kappa_2(j-k), 1-\xi\} + (1-p)\kappa_2\,\mathrm{IB}\{H_L, H_U; \kappa_1 k+\kappa_2(j-k+1), 1-\xi\}.$$

For Model (iii) with $G(v) = 1-Q_\delta\{(1-v)^\delta\}$, one has for any nonnegative integer s

$$\begin{aligned}
E_s &= \mathbb{E}\left\{(1-V)^{-\xi} Q_\delta\{(1-V)^\delta\}^s | H_L < V < H_U\right\} \\
&= \left(\frac{1+\delta}{\delta}\right)^s \mathbb{E}\left\{(1-V)^{s-\xi}\left(1-\frac{(1-V)^\delta}{1+\delta}\right)^s | H_L < V < H_U\right\} \\
&= \frac{1}{\delta^s}\sum_{j=0}^{s}\binom{s}{j}(-1)^j(1+\delta)^{s-j}\mathbb{E}\left\{(1-V)^{s-\xi+\delta j} | H_L < V < H_U\right\}. \\
&= \frac{1}{\delta^{s+1}\Delta_F}\sum_{j=0}^{s}\binom{s}{j}(-1)^j(1+\delta)^{s-j+1}\frac{B_j}{(s-\xi+\delta j+1)(s-\xi+\delta(j+1)+1)},
\end{aligned}$$

where the coefficients B_j are available in closed form as

$$\begin{aligned}
B_j = {} & (s-\xi+\delta j+1)\left\{(1-H_U)^{s-\xi+\delta(j+1)+1} - (1-H_L)^{s-\xi+\delta(j+1)+1}\right\} \\
& -(s-\xi+\delta(j+1)+1)\left\{(1-H_U)^{s-\xi+\delta j+1} - (1-H_L)^{s-\xi+\delta j+1}\right\}.
\end{aligned}$$

In particular, for noncensored PWMs ($H_L = 0$, $H_U = 1$), one has $B_j = \delta$. Thus, one gets

$$\mu_0 = \frac{\sigma}{1-\xi}\times\frac{2+\delta-\xi}{1+\delta-\xi},$$

$$\mu_1 = \frac{\sigma}{2(2-\xi)}\times\left\{\frac{2(1+\delta)^2}{\xi\delta(2+\delta-\xi)} - \frac{2(2-\xi)(1+\delta)}{\xi\delta(2+\delta-\xi)(2+2\delta-\xi)} - \frac{2-\xi}{\xi}\right\},$$

$$\mu_2 = \frac{\sigma}{3(3-\xi)}\times\left\{\frac{3(1+\delta)^3}{\xi\delta^2(3+\delta-\xi)} - \frac{6(3-\xi)(1+\delta)^2}{\xi\delta^2(3+\delta-\xi)(3+2\delta-\xi)} + \frac{3(3-\xi)(1+\delta)}{\xi\delta^2(3+2\delta-\xi)(3+3\delta-\xi)} - \frac{(3-\xi)}{\xi}\right\},$$

which can be compared to (B1). As expected, one recovers (B1) as $\delta \to \infty$.

Acknowledgments

Part of this work has been supported by the ANR-DADA, LEFE-INSU-Multirisk, AMERISKA, A2C2, CHAVANA and Extremoscope projects. The authors acknowledge Meteo France for the Lyon precipitation time series that available to anyone upon request. Part of the work was done when the first author was visiting the IMAGE-NCAR group in Boulder, CO, USA. The authors would also like very much to credit the contributors of the *R Core Team* [2013]. The data are freely available by sending an email to Philippe Naveau (naveau@lsce.ipsl.fr).

References

Ailliot, P., D. Allard, V. Monbet, and P. Naveau (2015), Stochastic weather generators: An overview of weather type models, *J. Soc. Française Stat.*, *156*, 101–113.

Azzalini, A. A. (1985), A class of distributions which include the normal, *Scand. J. Stat.*, *12*, 171–178.

Beirlant, J., E. Joossens, and J. Segers (2009), Second-order refined peaks-over-threshold modelling for heavy-tailed distributions, *J. Stat. Plann. Inference*, *139*, 2800–2815.

Caeiro, F., and M. I. Gomes (2011), Semi-parametric tail inference through probability-weighted moments, *J. Stat. Plann. Inference*, *141*(2), 937–950.

Carreau, J., and Y. Bengio (2009), A hybrid Pareto model for asymmetric fat-tailed data: The univariate case, *Extremes*, *12*, 53–76, doi: 10.1007/s10687-008-0068-0.

Carreau, J., and M. Vrac (2011), Stochastic downscaling of precipitation with neural network conditional mixture models, *Water Resour. Res.*, *47*, W10502, doi:10.1029/2010WR010128.

Carreau, J., P. Naveau, and E. Sauquet (2009), A statistical rainfall-runoff mixture model with heavy-tailed components, *Water Resour. Res.*, *45*, W10437, doi:10.1029/2009WR007880.

Chaussé, P. (2010), Computing generalized method of moments and generalized empirical likelihood with R, *J. Stat. Software*, *34-11*, 1–35.

Coles, S. G. (2001), *An Introduction to Statistical Modelling of Extreme Values*, *Springer Ser. Stat*.

Cooley, D., D. Nychka, and P. Naveau (2007), Bayesian spatial modeling of extreme precipitation return levels, *J. Am. Stat. Assoc.*, *102*(479), 824–840.

de Haan, L., and A. Ferreira (2006), *Extreme Value Theory: An Introduction*, *Springer Ser. Oper. Res. Finan. Eng*.

Deidda, R. (2010), A multiple threshold method for fitting the generalized Pareto distribution to rainfall time series, *Hydrol. Earth Syst. Sci.*, *14*, 2559–2575.

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

Diebolt, J., A. Guillou, P. Naveau, and P. Ribereau (2008), Improving probability-weighted moment methods for the generalized extreme value distribution, *REVSTAT Stat. J.*, *6*, 33–50.
Dupuis, D. J. (1999), Exceedances over high thresholds: A guide to threshold selection, *Extremes*, *1*(3), 251–261.
Embrechts, P., C. Klüppelberg, and T. Mikosch (1997), *Modelling Extremal Events for Insurance and Finance, Volume 33 of Applications of Mathematics*, Springer, Berlin.
Falk, M., J. Hüsler, and R.-D. Reiss (2010), *Laws of Small Numbers: Extremes and Rare Events*, 3rd ed., Birkhäuser Verlag, Basel, Switzerland.
Ferreira, A., and L. de Haan (2014), On the block maxima method in extreme value theory. [Available at http://arxiv.org/abs/1310.3222.]
Fisher, R. A., and L. H. C. Tippett (1928), Limiting forms of the frequency distribution of the largest or smallest member of a sample, *Proc. Cambridge Philos. Soc.*, *24*, 180–190.
Frigessi, A., O. Haug, and H. Rue (2002), A dynamic mixture model for unsupervised tail estimation without threshold selection, *Extremes*, *5*, 219–235.
Genton, M. G. (2004), *Skew-Elliptical Distributions and Their Applications: A Journey Beyond Normality*, edited volume, 416 pp., Chapman and Hall, Boca Raton, Fla.
Hosking, J. R. M., and J. R. Wallis (1987), Parameter and quantile estimation for the generalized Pareto distribution, *Technometrics*, *29*, 339–349.
Kallache, M., M. Vrac, P. Naveau, and P. A. Michelangeli (2011), Nonstationary probabilistic downscaling of extreme precipitation, *J. Geophys. Res.*, *116*, D05113, doi:10.1029/2010JD014892.
Katz, R., M. Parlange, and P. Naveau (2002), Extremes in hydrology, *Adv. Water Resour.*, *25*, 1287–1304.
Katz, R. W. (1977), Precipitation as a chain-dependent process, *J. Appl. Meteorol.*, *16*, 671–676.
Koch, E., and P. Naveau (2015), A frailty-contagion model for multi-site hourly precipitation driven by atmospheric covariates, *Adv. Water Resour.*, *78*, 145–154.
Landwehr, J., N. Matalas, and J. R. Wallis (1979), Probability weighted moments compared with some traditional techniques in estimating Gumbel parameters and quantiles, *Water Resour. Res.*, *15*, 1055–1064.
Li, C., V. P. Singh, and A. K. Mishra (2012), Simulation of the entire range of daily precipitation using a hybrid probability distribution, *Water Resour. Res.*, *48*, W03521, doi:10.1029/2011WR011446.
MacDonald, A., C. Scarrott, D. Lee, B. Darlow, M. Reale, and G. Russell (2011), A flexible extreme value mixture model, *Comput. Stat. Data Anal.*, *55*, 2137–2157.
Naveau P., A. Toreti, I. Smith, and E. Xoplaki (2014), A fast nonparametric spatio-temporal regression scheme for generalized Pareto distributed heavy precipitation, *Water Resour. Res.*, *50*, 4011–4017, doi:10.1002/2014WR015431.
Papastathopoulos, I., and A. J. Tawn (2013), Extended generalised Pareto models for tail estimation, *J. Stat. Plann. Inference*, *143*, 131–143.
R Core Team (2013), *R: A Language and Environment for Statistical Computing*, R Found. for Stat. Comput., Vienna. [Available at http://www.R-project.org/.]
Ribereau, P., E. Masiello, and P. Naveau (2015), Skew generalized extreme value distribution: Probability weighted moments estimation and application to block maxima procedure, in *Communications in Statistics—Theory and Methods*, doi:10.1080/03610926.2014.935434. [Available at http://www.tandfonline.com/action/doSearch?quickLinkJournal=&journalText=&AllField=ribereau&publication=40000844.]
Robert, C., and G. Casella (2004), *Monte Carlo Statistical Methods*, Springer Texts Stat.
Sartori, N. (2006), Bias prevention of maximum likelihood estimates for scalar skew normal and skew t distributions, *J. Stat. Plann. Inference*, *136*, 4259–4275.
Tancredi, A., C. Anderson, and A. O'Hagan (2006), Accounting for threshold uncertainty in extreme value estimation, *Extremes*, *9*, 87–106.
Vrac, M., and P. Naveau (2007), Stochastic downscaling of precipitation: From dry events to heavy rainfalls, *Water Resour. Res.*, *43*, W07402, doi:10.1029/2006WR005308.
Vrac, M., M. Stein, and K. Hayhoe (2007), Statistical downscaling of precipitation through a nonhomogeneous stochastic weather typing approach, *Clim. Res.*, *34*, 169–184, doi:10.3354/cr00696.
Wilks, D. (2006), *Statistical Methods in the Atmospheric Sciences*, 2nd ed., Elsevier, Oxford, U. K.