Flexible semiparametric Generalized Pareto modeling of the entire range of rainfall amount

Patricia Tencaliec, Anne-Catherine Favre, Philippe Naveau, Clémentine Prieur

▶ **To cite this version:**

Patricia Tencaliec, Anne-Catherine Favre, Philippe Naveau, Clémentine Prieur. Flexible semiparametric Generalized Pareto modeling of the entire range of rainfall amount. Environmetrics, inPress, pp.1-22. hal-01709061v1

HAL Id: hal-01709061

https://inria.hal.science/hal-01709061v1

Submitted on 14 Feb 2018 (v1), last revised 9 Jul 2019 (v2)

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Flexible semiparametric Generalized Pareto modeling of the entire range of rainfall amount

Tencaliec, P.[1], Favre, A.-C.[2], Naveau, P.[3], and Prieur, C.[1]

[1]Univ. Grenoble Alpes, CNRS, INRIA, LJK, F-38000 Grenoble, France
[2]Univ. Grenoble Alpes, CNRS, IRD, Grenoble INP, IGE, F-38000 Grenoble, France
[3]Laboratoire des Sciences du Climat et de l'Environnement, LSCE/IPSL, CNRS-CEA-UVSQ, Université Paris-Saclay, Gif-sur-Yvette, France

Abstract

Precipitation amounts at daily or hourly scales are skewed to the right and heavy rainfall is poorly modeled by a simple gamma distribution. An important, yet challenging topic in hydrometeorology is to find a probability distribution that is able to model well low, moderate and heavy rainfall. To address this issue, we present a semiparametric distribution suitable for modeling the entire-range of rainfall amount. This model is based on a recent parametric statistical model called the class of Extended Generalized Pareto Distributions (EGPD). The EGPD family is in compliance with Extreme Value Theory for both small and large values, while it keeps a smooth transition between these tails and bypasses the hurdle of selecting thresholds to define extremes. In particular, return levels beyond the largest observation can be inferred. To add flexibility to this EGPD class, we propose to model the transition function in a nonparametric fashion. A fast and efficient nonparametric scheme based on Bernstein polynomial approximations is investigated. We perform simulation studies to assess the performance of our approach. It is compared to two parametric models: a parametric EGPD and the classical Generalized Pareto Distribution (GPD), the later being only fitted to excesses above high threshold. We also apply our semiparametric version of EGPD to daily rainfall data recorded at Mont-Aigoual weather station in France.

Keywords— precipitation, Extreme Value Theory, Extended Generalized Pareto Distribution, semiparametric, Bernstein polynomials, maximum likelihood estimator

1 Introduction

Modeling the distribution of precipitation data is needed in many applications regarding water resources management, design, or planning, such as urban water supplies, hydropower, forecast of flood or droughts events, irrigation systems. A first and essential step in the statistical modeling is to find probability distributions that can describe correctly the occurrences and the intensities of precipitation. As the process of rainfall occurrences is discrete, while its amount is a continuous one, the most common approach is to have a different model for these two features. In this work, we only focus on the second part, *i.e.,* the statistical modeling of strictly positive rainfall amounts, and we refer to Wilks (1999) and Apipattanavis *et al.* (2007) to model occurrence processes.

Fitting accurately the full spectrum of rainfall amounts has proven to be a challenging task, mainly due to the fact that they are heavily skewed to the right. Different distributions, such as gamma (see *e.g.*, Katz, 1977; Stern and Coe, 1984; Wilks, 1989), exponential (see *e.g.*, Woolhiser and Pegram, 1979; Richardson, 1981; Wilks, 1999; Garavaglia *et al.*, 2010), Weibull (see *e.g.*, Zucchini and Adamson, 1984) or lognormal (see *e.g.*, Apipattanavis *et al.*, 2007) have been considered as possible candidates. As suggested by Vrac *et al.* (2007) and Wilks (2011), gamma and mixed exponential are typically the preferred choices, but, as pointed out by Katz *et al.* (2002), the tail of a gamma distribution can be too light to model heavy rainfall and underestimation of extreme values can occur, an undesirable feature in any hydrological risk analysis.

As mentioned by Evin *et al.* (2018), stochastic precipitation generators have become useful tools in risk assessment studies for two reasons. Realistic simulated precipitation draws are needed as inputs of conceptual hydrological models. In particular, the observed series of streamflows are too short to estimate the very high floods return levels. In this context, simple but rich probability density functions (pdf) to generate precipitation draws, extreme included, are needed. In this framework, the work in this paper can also be viewed as proposing a new and flexible tool (the precipitation building unit) to researchers interested by constructing such stochastic rainfall weather generators.

As the upper tail of the distribution holds crucial information, the Generalized Pareto Distribution (GPD) is nowadays the common choice for modeling heavy rainfall in the statistical climatological community (see, *e.g.*, Katz *et al.*, 2002; Nadarajah, 2005). GPD is defined by the cumulative distribution function (cdf) $H_\xi(x/\sigma)$ as

$$H_\xi(z) = \begin{cases} 1 - (1+\xi z)_+^{-1/\xi}, & \text{for } \xi \neq 0, \\ 1 - e^{-z}, & \text{for } \xi = 0, \end{cases}$$

where ξ is the shape parameter, $\sigma > 0$ is the scale parameter and $a_+ = \max(a, 0)$. It is mathematically justified by extreme value theory (EVT) (see, *e.g.*, Coles, 2001). In hydrology, ξ is often assumed to be non negative for daily rainfall (see, *e.g.*, Evin *et al.*, 2018). We keep this hypothesis of $\xi \geq 0$ in this work.

A practical limitation of the GPD is that it can be only applied to "extreme precipitation" and this leads to the question of how to set a threshold that differentiates heavy and moderate rainfall. Answering this question becomes delicate when the number of time series under study increases, say in a climate model output analysis with thousands of grid points. In such situations, graphical device tools like a Quantile-Quantile plot (QQ-plot) (see, *e.g.*, Coles, 2001; Katz *et al.*, 2002) cannot be visually checked anymore. Hence, the threshold in hydrological instances dealing with a large number of datasets is chosen arbitrarily, classically the 95% quantile of each time series.

In addition, practitioners can be interested in summarizing the full rainfall range and not only the extremal behavior, *e.g.*, to determine in a climate change Detection & Attribution study if rainfall (extremes included) have changed over time (see, *e.g.*, Hegerl and Zwiers, 2011). In recent years, a few attempts have been made to bypass the threshold choice and to characterize the full precipitation spectrum. Mixture and hybrid models have been proposed, such as the dynamic weighted model used by Vrac and Naveau (2007), or the hybrid model based on a mixture of gamma and GPD distributions of Furrer and Katz (2008). In practice, these models have a large number of parameters and the inference remains a challenge (see also Carreau and Bengio, 2008; MacDonald *et al.*, 2011, for mixture approaches).

Moving away from the idea of a mixture, Naveau *et al.* (2016) recently proposed a construction that allows a smooth transition between GPD type tails and the middle part (bulk) of the distribution. Here, this class of models is referred as the Extended Generalized Pareto Distribution (EGPD) family. It bypasses the thresholds selection step and it is in compliance with EVT, not only for heavy rainfall, but also for low precipitation amounts. In particular, low precipitation amounts are classically modeled by a gamma distribution.

Mathematically, a member of the EGPD model has to be expressed as

$$F(x) = G\left\{H_\xi(x/\sigma)\right\}, \text{ for all } x > 0, \tag{1}$$

or, in terms of densities, as $f(x) = \frac{1}{\sigma} g\left\{H_\xi(x/\sigma)\right\} \cdot h_\xi(x/\sigma)$, where h_ξ and H_ξ represent the pdf and the cdf of the GPD, while g (*resp.* G) denotes a continuous pdf (*resp.* cdf) on the unit interval. To insure that the upper tail behavior of F is driven by a GPD with parameter ξ, the survival function $\bar{G} = 1-G$ has to satisfy that the limit $a := \lim_{u \downarrow 0} \frac{\bar{G}(1-u)}{u}$ with $u = \overline{H}_\xi(\frac{x}{\sigma})$ is finite and positive as x tends to infinity (u tends to 0). In this case, the upper tail behavior of $\overline{F}(x)$ is equivalent to the original GPD tail used to build $F(x)$, *i.e.*, $\frac{\overline{F}(x)}{\overline{H}_\xi(\frac{x}{\sigma})} = \frac{\bar{G}(1-u)}{u} \sim a > 0$ as x tends to $+\infty$. To force low rainfall (modeled as $-X$) to follow a GPD for small values near zero, we need that the limit $c := \lim_{u \downarrow 0} \frac{G(u)}{u^s}$ is positive and finite for some positive real s. In this case, $F(x) \sim c\left\{H_\xi(\frac{x}{\sigma})\right\}^s \sim \frac{c}{\sigma^s} x^s$ as x tends to zero.

In Naveau *et al.* (2016), four parametric models for the G function satisfying the required constraints were proposed and compared. But, besides inferential

convenience, there is not a theoretical reason to choose a particular parametric G. In this context, the main goal of the present work is to determine if a nonparametric family for G in Equation (1) can be proposed, and more importantly, if this model can be quickly and efficiently inferred. This will bring flexibility to this family and it will be a versatile tool for hydrologists. To reach this target, we take advantage of Bernstein polynomials approximation by relying on the link between Bernstein polynomials and the beta distributions.

The paper is organized as follows. In Section 2, a short background on Bernstein polynomials and its relationship with density estimation is provided. Our proposed semiparametric EGPD model is also described in Section 2 . Section 3 discusses in depth the estimation procedure. Sections 4 and 5 are dedicated to case studies on simulated and rainfall datasets, respectively. Conclusions and perspectives are presented in Section 6.

2 semiparametric EGPD model class

Naveau *et al.* (2016) explained that the key component of the EGPD class is the continuous function G on $[0, 1]$ (called transition function). This building block links the bulk of the distribution with both the upper and lower tails. As our goal is to propose a flexible form for G, a first idea could be to work with Gaussian mixtures. But, as the support of G is the unit interval, this option will lead to overlay complex truncations and boundaries effects. Another alternative could be mixtures of beta densities with a few components, *e.g.*, Ji *et al.* (2005). However, parametric mixture models become problematic when the number of components increases, because too many parameters have to be estimated. Another issue is that observed rainfall are measured in millimeters, but G describes data on $[0, 1]$. This implies that the estimation of G is based on pseudo-observations $H_\xi(X_i/\sigma)$, thus requiring as a first step the estimation of $H_\xi(x/\sigma)$.

This implies that any nonparametric estimation of G has to be simple in order to keep at bay computational issues. In this context, kernel density estimators, polynomials approximation, or projections techniques, rather than mixture models, should be favored. A natural way to approximate functions on the unit interval is to use Bernstein polynomials.

2.1 Bernstein polynomials and density estimation

In 1912, Bernstein introduced the polynomials named after him in a proof of the Weierstrass Approximation Theorem. He showed that any continuous function G on the interval $[0, 1]$ can be approximated up to some degree of accuracy by Bernstein polynomials. Hence, if G denotes a continuous cdf on $[0, 1]$, it can be approximated by the Bernstein estimator of degree $m > 0$ defined by

$$P_m(t, G) = \sum_{k=0}^{m} G\left(\frac{k}{m}\right) b_{k,m}(t),$$

where $b_{k,m}(t) = \binom{m}{k} t^k (1-t)^{m-k}$ for $t \in [0,1]$. These so-called Bernstein bases have attractive properties, *e.g.*, non-negativity, partition of unity and symmetry (see Farouki, 2012, for details). To use the approximation $P_m(t, G)$ in a statistical context, one needs to compute $G\left(\frac{k}{m}\right)$ from a sample drawn from G. The idea of Vitale (1975) was to estimate these coefficients by replacing the unknown G by its empirical cumulative distribution function (ecdf), say $G_n(t)$. This strategy led Vitale to propose the following estimator of the pdf $g(t)$

$$\hat{g}_{m,n}(t) = m \sum_{k=0}^{m-1} \left\{ G_n\left(\frac{k+1}{m}\right) - G_n\left(\frac{k}{m}\right) \right\} b_{k,m-1}(t) \tag{2}$$

that is a valid density (positive and $\int_0^1 \hat{g}_{m,n}(t)dt = 1$) because of $G_n(0) = 0$ and $G_n(1) = 1$. Babu *et al.* (2002) showed that the degree m should be chosen such that $m \in \{2, \ldots, [n/\log(n)]\}$ for large samples of size n. More recently, Leblanc (2010, 2012a,b) studied the boundary properties of both density and distribution estimators, (see also Petrone, 1999; Ghosal, 2001; Kakizawa, 2004; Bouezmarni and Rolin, 2007, for extensions in a Bayesian and/or multivariate context).

As mentioned by Vitale (1975), all approximations based on $b_{k,m}(t)$ can be rewritten in terms of linear combinations of beta densities. In particular, we rewrite $\hat{g}_{m,n}(t)$ as a sum of beta densities with the following notation

$$\hat{g}_{m,n}(t) = \sum_{k=1}^{m} \omega_{k,m} \beta_{k,m-k+1}(t), \tag{3}$$

where $t \in [0,1]$, $\omega_{k,m} = G_n\left(\frac{k}{m}\right) - G_n\left(\frac{k-1}{m}\right)$, and $\beta_{a,b}(t) = t^{a-1}(1-t)^{b-1}/B(a,b)$ corresponds to the classical beta pdf with parameters a and b, respectively, and $B(a,b)$ denotes the beta function. Here, we stress that, given m, the approximation $\hat{g}_{m,n}(t)$ is not a classical mixture of densities with unknown parameters. The beta coefficients a and b are known ($a = k$ and $b = m - k + 1$) and the weights are straightforward to compute if G_n is given. In other words, although $\hat{g}_{m,n}(t)$ is a mixture of beta pdfs, it has to be interpreted as an expansion on the beta "bases". This implies that, given G_n, the only unknown is m. It can be large and interpreted as a type of bandwidth. Finding m corresponds to resolving a bias-variance tradeoff (see *e.g.*, Vitale, 1975; Leblanc, 2012b).

2.2 EGPD model class based on Bernstein-beta density

Now, the EGPD family defined by Equation (1) and the Bernstein approximation captured by Equation (3) can be combined via

$$F_{m,n,\boldsymbol{\theta}}(x) = \hat{G}_{m,n}\left\{H_\xi(x/\sigma)\right\}, \tag{4}$$

where $\hat{G}_{m,n}(t)$ represents the cdf associated with $\hat{g}_{m,n}(t)$ – see Equation (3) – and $\boldsymbol{\theta} = (\sigma, \xi)^t$ corresponds to the GPD parameters. At this stage, we need to determine the constraints on $\hat{G}_{m,n}$ in order that F belongs to the EGDP class.

Lemma 1 *If among all the coefficients* $\omega_{k,m} = G_n\left(\frac{k}{m}\right) - G_n\left(\frac{k-1}{m}\right)$ *with* $k = 1, \ldots, m$, *the last one is positive, i.e.,* $\omega_{m,m} > 0$, *then we have*

1. $\lim_{x\to 0} \frac{F_{m,n,\boldsymbol{\theta}}(x)}{x^s} = \frac{m}{s}\sigma^{-s}\binom{m-1}{s-1}\omega_{s,m} > 0$, *where* s *denotes the position of the first non-null weight in* $\boldsymbol{\omega} = (\omega_{1,m}, \ldots, \omega_{m,m})^t$;
2. $\lim_{x\to\infty} \frac{\overline{F}_{m,n,\boldsymbol{\theta}}(x)}{\overline{H}_\xi(x/\sigma)} = m\omega_{m,m} > 0$.

In addition, let Y *be any non-negative continuous random variable such that the conditional limit* $\mathbb{P}(Y > x + u | Y > u)$ *goes, as* u *gets large, to* $(1 + \tilde{\xi}\frac{x}{\tilde{\sigma}})^{-1/\tilde{\xi}}$ *for some parameters* $\tilde{\sigma}$ *and* $\tilde{\xi} > 0$.

3. *If* Y *can be rewritten as* $Y = \sigma H_\xi^{-1}\left\{G^{-1}(U)\right\}$ *with the survival of the cdf* G *satisfying* $\lim_{w\downarrow 0} \frac{\overline{G}(1-w)}{w} \in (0, \infty)$, *then* $\tilde{\xi} = \xi$.

[Proof of Lemma 1] The proof is postponed to the appendix.

Remark 2 *From Item 1. of Lemma 1 above, we conclude that the lower tail behavior of the model described by Equation* (4) *is controlled by the rank of the first non-null weight in* $\boldsymbol{\omega}$. *From Item 2. of Lemma 1, we see that the assumption* $\omega_{m,m} > 0$ *is required to prove that the upper tail behavior in our model is described by a GPD. Item 3. tells us that imposing* $\lim_{w\downarrow 0} \frac{\overline{G}(1-w)}{w} \in (0, \infty)$ *insures that our* EGPD *tail behavior driven by* ξ *is equivalent to the one obtained using a genuine* EVT *argument. This explains why our Bernstein approximation of* G *has to satisfy this condition too. As*

$$\lim_{w\to 0} \frac{\overline{\hat{G}}_{m,n}(1-w)}{w} = m\omega_{m,m},$$

combining Item 2 and Item 3 implies that the constraint $\omega_{m,m} > 0$ *is sufficient to make sure that our Bernstein approximation,* $\hat{G}_{m,n}$, *does not impact the expected upper GPD tail behavior (under the condition that the observations can be rewritten as* $\sigma H_\xi^{-1}\left\{G^{-1}(U)\right\}$*).*

In practice, we have to deal with a few issues. First, we do not observe random draws from G, but only rainfall measurements from F. Hence, we do not have a direct way to compute the ecdf G_n, and consequently the weights $\omega_{k,m} = G_n\left(\frac{k}{m}\right) - G_n\left(\frac{k-1}{m}\right)$. In addition, we have the constraint $\omega_{m,m} > 0$ that may not be always satisfied. Last but not least, the number of components m in $\hat{g}_{m,n}(t)$ has to be chosen and the parameters $\boldsymbol{\theta} = (\sigma, \xi)^t$ have to be inferred. On the positive side, we can notice that, if random draws from G were available and m given, then the weights $\omega_{k,m}$ can be instantaneously computed, *i.e.*, there is no need of running a time consuming optimization scheme in such a case. In this context, our inferential strategy is to use a recursive argument. Basically, we infer the hidden values drawn from G, compute and adjust the weights, and then estimate $\boldsymbol{\theta} = (\sigma, \xi)^t$. This algorithm can be repeated until the values of some criterion are stable. The next section will detail our strategy.

3 Methodology for fitting a semiparametric EGPD model

3.1 Initial values

For rainfall data, Naveau *et al.* (2016) noticed that the special case of $G(u) = u^\kappa$ with $\kappa > 0$ provided a decent fit for hourly and daily precipitation in France. Evin *et al.* (2018) used this same G to model precipitation in Switzerland. Consequently, this parametric model for G appears to be a good starting point to give initial estimates for $\boldsymbol{\theta} = (\sigma, \xi)^t$. Let us call them $\hat{\boldsymbol{\theta}}_0 = (\hat{\sigma}_0, \hat{\xi}_0)^t$. From Equation (1), it is possible to show that the random variable $H_\xi\left(X/\sigma\right)$ follows the cdf G if X follows F. Suppose that $(X_1, \ldots, X_n)$ represents our observed rainfall sample with cdf F, then we can introduce the random variables

$$\hat{Z}_i = H_{\hat{\xi}_0}\left(\frac{X_i}{\hat{\sigma}_0}\right), \text{ for } i = 1, \ldots, n. \tag{5}$$

These reconstructed random variables should mimic the hidden sample driven by G. Consequently, our initial approximation of the weights $\omega_{k,m}$, for a given m, is

$$\hat{\omega}_{k,m} = \hat{G}_n\left(\frac{k}{m}\right) - \hat{G}_n\left(\frac{k-1}{m}\right)$$

where $\hat{G}_n$ represents the ecdf obtained from $(\hat{Z}_1, \ldots, \hat{Z}_n)$. To complete our initialization process, we need to force the last weight $\hat{\omega}_{m,m}$ to be non null. If $\hat{\omega}_{m,m}$ is non null, then we are done. Otherwise, we take $\hat{\omega}_{m,m} = \hat{\kappa}_0/m$ and we renormalize the weights to insure that $\hat{\omega}_{1,m} + \cdots + \hat{\omega}_{m,m} = 1$ (see Algorithm 1 for more details). This last choice is motivated by the fact that, if $G(u) = u^\kappa$, parameter $\omega_{m,m}$ is asymptotically equivalent to $\frac{\kappa}{m}$ as m goes to infinity.

Algorithm 1 Weights approximation given $\mathbf{x} = (x_1, \ldots, x_n)^t, \sigma, \xi, \kappa, m$

procedure INPUTS:$(\mathbf{x}, \sigma, \xi, \kappa, m)$
  $n = \text{length}(x)$
  $z_i = H_\xi(x_i/\sigma)$
  for each integer k in $1 : m$ **do**
    $\omega_{k,m} = G_n(\frac{k}{m}) - G_n(\frac{k-1}{m})$ with G_n ecdf obtained from $(z_1, \ldots, z_n)^t$
    if ($k = m$ and $\omega_{k,m} = 0$) **then**
      $\omega_{k,m} = \frac{\kappa}{m}$
      $\boldsymbol{\omega} = \boldsymbol{\omega}/\sum\boldsymbol{\omega}$ ▷ normalization to add to 1
    end if
  end for
  Return $\boldsymbol{\omega}$
end procedure

3.2 Estimation of $\boldsymbol{\theta} = (\sigma, \xi)^t$ via likelihood maximization

For this step, we assume that m and the weights are given and the only unknowns are the two parameters $\boldsymbol{\theta} = (\sigma, \xi)^t$. Maximum likelihood estimator (MLE) is a standard inferential technique in statistics, (*e.g.*, see Davison, 1984; Grimshaw, 1993, for its application to the GPD). In our context, we want to estimate the parameters $\boldsymbol{\theta} = (\sigma, \xi)$ in

$$\hat{f}_{m,n,\boldsymbol{\theta}}(x) = \frac{1}{\sigma}\hat{g}_{m,n}\left\{H_\xi(x/\sigma)\right\} h_\xi(x/\sigma).$$

More specifically, the optimization problem that must be solved is

$$\begin{aligned} \hat{\boldsymbol{\theta}}^{\text{MLE}} &= \text{argmin}\left\{\text{-log}\mathcal{L}(\boldsymbol{\theta}|\mathbf{x})\right\} \\ &\quad \text{with } \boldsymbol{\theta} = (\sigma, \xi)^t \text{ subject to } \quad \xi \geq 0 \text{ and } \sigma > 0, \end{aligned} \tag{6}$$

where

$$\log\mathcal{L}(\boldsymbol{\theta}|\mathbf{x}) = \sum_{i=1}^{n} \log\left[\hat{g}_{m,n}\left\{H_\xi(x_i/\sigma)\right\}\right] + \sum_{i=1}^{n} \log\left\{h_\xi(x_i/\sigma)\right\} - n\log\sigma$$

represents the log-likelihood. To avoid the inequality constraints, $\xi \geq 0$ and $\sigma > 0$, the optimization is done with the re-parametrization $\sigma = \exp(\sigma^R)$ and $\xi = \left(\xi^R\right)^2$, for $(\sigma^R, \xi^R) \in \mathbb{R} \times \mathbb{R}_+$. Still, we cannot obtain an analytic form solution of Equation (6), thus, the optimal solution $\hat{\boldsymbol{\theta}}^{\text{MLE}}$ is computed numerically through non-linear optimization solvers. From the MLE of $\boldsymbol{\theta}$, we can come back to Equation (5) and update the estimated weights $\hat{\omega}_{k,m}$, and then restart a new estimation of $\boldsymbol{\theta}$. Algorithm 2 in the next section summarizes the details of such a loop.

3.3 Main algorithm

Algorithm 2 Estimation of $(\omega_{1,m}, \ldots, \omega_{m,m})^t$, σ and ξ in Equation (4) for a given m

procedure INPUTS($\mathbf{x}$ and m)
  $\sigma_0, \xi_0, \kappa_0$ ▷ fit a EGDP to $\mathbf{x}$ with $G(u) = u^\kappa$;
  $\boldsymbol{\omega}_0 =$ ALGORITHM 1$(\mathbf{x}, \sigma_0, \xi_0, \kappa_0, m)$ ▷ initialize weights
  $\boldsymbol{\theta}^{\text{MLE}}_{init} =$ argmin $\{\text{-log}\mathcal{L}(\boldsymbol{\theta}|\mathbf{x}, \boldsymbol{\omega}_0)\}$ ▷ estimate σ and ξ
  $\boldsymbol{\omega}_{init} =$ ALGORITHM 1$(\mathbf{x}, \sigma^{\text{MLE}}_{init}, \xi^{\text{MLE}}_{init}, \kappa_0, m)$ ▷ new weights
  $cond = true, eps = 10^{-6}$
  while $cond = true$ **do**
    $\boldsymbol{\theta}^{\text{MLE}}_{new} =$ argmin $\{\text{-log}\mathcal{L}(\boldsymbol{\theta}|\mathbf{x}, \boldsymbol{\omega}_{init})\}$ ▷ estimate σ and ξ
    $\boldsymbol{\omega}_{new} =$ ALGORITHM 1$(\mathbf{x}, \sigma^{\text{MLE}}_{new}, \xi^{\text{MLE}}_{new}, \kappa_0, m)$ ▷ ω for new parameters
    if $(\log\mathcal{L}(\boldsymbol{\theta}^{\text{MLE}}_{new}|\mathbf{x}) - \log\mathcal{L}(\boldsymbol{\theta}^{\text{MLE}}_{init}|\mathbf{x})) < eps$ **then**
      $cond = false$
    end if
    $\log\mathcal{L}(\boldsymbol{\theta}^{\text{MLE}}_{init}|\mathbf{x}) = \log\mathcal{L}(\boldsymbol{\theta}^{\text{MLE}}_{new}|\mathbf{x}); \boldsymbol{\theta}^{\text{MLE}}_{init} = \boldsymbol{\theta}^{\text{MLE}}_{new}; \boldsymbol{\omega}_{init} = \boldsymbol{\omega}_{new}$
  end while
  Return $\boldsymbol{\theta}^{\text{MLE}}_{new}, \boldsymbol{\omega}_{new}$
end procedure

Remark 3 *Algorithm 2 has some similarities with EM algorithms (see, e.g., McLachlan and Krishnan (2007) and references therein).The density in our model has the form*

$$f_{m,\boldsymbol{\theta}}(x) = \frac{1}{\sigma} g_m \left\{H_\xi(x/\sigma)\right\} h_\xi(x/\sigma)$$

with $g_m(t) = \sum_{k=1}^m \omega_{k,m} \beta_{k,m-k+1}(t)$. *As mentioned in Section 2.1,* $f_{m,\boldsymbol{\theta}}(x)$ *is not an usual mixture of densities, as the weights* $w_{k,m}$, $k = 1, \ldots, m$ *have a particular interpretation,* i.e., $w_{k,m} = G(\frac{k}{m}) - G(\frac{k-1}{m})$. *However, it is possible to introduce the discrete random variable* W *with possible values in the set* $\{1, \ldots m\}$ *and such that for* $k = 1, \ldots, m$, *the probability* $\mathbb{P}(W = k) = w_{k,m}$ *corresponds to a choice of a particular weight. It is then possible to interpret each step of our algorithm as follows. Algorithm 1 can be considered as the Expectation step. Let us define the parameter* $\psi := (\sigma, \xi, \kappa)$. *In the initialization of the algorithm, we start from a first guess* ψ_0 *obtained from the EGPD model. Parameter* κ *is fixed once for all to the value* κ_0. *We then focus on the update of parameter* $\theta = (\sigma, \xi)$. *Let* $\log\mathcal{L}(\mathbf{X}, W|\theta)$ *the log-likelihood of complete data. At Iteration* 1 *of the algorithm, we compute the conditional expectation* $\mathbb{E}_{W|\mathbf{X},\theta_0} \log\mathcal{L}(\mathbf{X}, W|\theta)$. *At Iteration* $p+1$, *we compute* $\mathbb{E}_{W|\mathbf{X},\theta_p} \log\mathcal{L}(\mathbf{X}, W|\theta)$,

where θ_p is the value of parameter θ obtained at iteration p of the Maximization step. We have

$$\log \mathcal{L}(\mathbf{X}, W|\theta) = \sum_{i=1}^{n}\sum_{k=1}^{m} \mathbb{1}_{W_i=k} \log\left[\frac{1}{\sigma}\beta_{k,m-k+1}\left\{H_\xi\left(\frac{X_i}{\sigma}\right)\right\} h_\xi\left(\frac{X_i}{\sigma}\right)\right].$$

Thus

$$\mathbb{E}_{W|\mathbf{X},\theta_p} \log \mathcal{L}(\mathbf{X}, W|\theta) = \sum_{i=1}^{n}\sum_{k=1}^{m} \mathbb{P}(W_i = k|X_i, \theta_p) \log\left[\frac{1}{\sigma}\beta_{k,m-k+1}\left\{H_\xi\left(\frac{X_i}{\sigma}\right)\right\} h_\xi\left(\frac{X_i}{\sigma}\right)\right]. \tag{7}$$

The particularity of our algorithm is that we estimate $\mathbb{P}(W_i = k|X_i, \theta_p)$ non-parametrically by

$$\frac{1}{n}\sum_{i=1}^{n} \mathbb{1}\left\{\frac{k-1}{m} \leq H_{\xi_p}\left(\frac{X_i}{\sigma_p}\right) \leq \frac{k}{m}\right\}. \tag{8}$$

Then, replacing $\mathbb{P}(W_i = k|X_i, \theta_p)$ by the estimate defined in Equation (8) in Equation (7), we get $\widehat{\mathbb{E}}_{W|\mathbf{X},\theta_p} \log \mathcal{L}(\mathbf{X}, W|\theta)$. Lines 6 to 9 of Algorithm 1 are specific to the constraint $w_{m,m} > 0$.

At Iteration p of the Maximization step (loop while in Algorithm 2 above), parameter $\theta = (\xi, \sigma)$ is updated by maximizing $\widehat{\mathbb{E}}_{W|\mathbf{X},\theta_p} \log \mathcal{L}(\mathbf{X}, W|\theta)$. Using classical arguments, it is possible to prove that

$$\Delta(\theta, \theta_p) := \log \mathcal{L}(\mathbf{X}|\theta) - \log \mathcal{L}(\mathbf{X}|\theta_p) \geq \sum_{k=1}^{m} \mathbb{P}(W = k|\mathbf{X}, \theta_p) \log\left(\frac{\mathcal{L}(\mathbf{X}, k|\theta)}{\mathcal{L}(\mathbf{X}, k|\theta_p)}\right) =: \delta(\theta|\theta_p).$$

We have $\widetilde{\theta}_{p+1} := \operatorname{argmax}_\theta \delta(\theta|\theta_p) = \operatorname{argmax}_\theta \mathbb{E}_{W|\mathbf{X},\theta_p} \log \mathcal{L}(\mathbf{X}, W|\theta)$ and $\delta(\theta_p|\theta_p) = 0$. Thus $\Delta(\widetilde{\theta}_{p+1}, \theta_p) := \log \mathcal{L}(\mathbf{X}|\widetilde{\theta}_{p+1}) - \log \mathcal{L}(\mathbf{X}|\theta_p) \geq 0$. As we maximize $\widehat{\mathbb{E}}_{W|\mathbf{X},\theta_p} \log \mathcal{L}_n(\mathbf{X}, W|\theta)$ instead of $\mathbb{E}_{W|\mathbf{X},\theta_p} \log \mathcal{L}(\mathbf{X}, W|\theta)$, it is not an easy task to study the increase of the likelihood nor the convergence for our algorithm. In practice, Algorithm 2 appears to perform adequately, see Section 4.

3.4 Selection of the Bernstein polynomial degree m

The number of components in the mixture, or alternatively the degree of the polynomial, is a very important feature of the model as it directly influences the smoothness of the estimator. Babu *et al.* (2002) showed that m should be of order $o\{n/\log(n)\}$ for consistent convergence results. Also, their numerical study indicated that the setting $m = \frac{n}{\log(n)}$ works well. But, they only covered small sample sizes (up to 125 observations), so when working with larger samples, the degree $m = \frac{n}{\log(n)}$ could be too large.

To avoid this issue, we prefer to rely on a data driven approach and use the Least Square Cross Validation (LSCV) scheme which is based on the minimization of the Mean Integrated Squared Error (MISE) (see Kakizawa, 2004;

Leblanc, 2010). The notation $\hat{f}^{(-i)}_{m,n,\theta}(x)$ and $\hat{g}^{(-i)}_{m,n}(t)$ indicate the same estimators based on all data but X_i, respectively. The optimal polynomial degree m (see Bouezmarni and Rolin, 2007, in the case of Bernstein estimators) is the integer that minimizes

$$\begin{aligned}\text{MISE}(m) =&\mathbb{E}\left[\int_0^\infty \left\{\hat{f}_{m,n,\theta}(x) - f(x)\right\}^2 \mathrm{d}x\right] \\ =&\mathbb{E}\left\{\int_0^\infty \hat{f}^2_{m,n,\theta}(x)\mathrm{d}x\right\} - 2\mathbb{E}\left\{\int_0^\infty \hat{f}_{m,n,\theta}(x)f(x)\mathrm{d}x\right\} + \int_0^\infty f^2(x)\mathrm{d}x.\end{aligned} \tag{9}$$

The last term in Equation (9) does not depend on m, thus it can be dropped to seek for the minimizer of MISE(m). We then search for the degree that minimizes the quantity $\text{MISE}(m) - \int_0^\infty f^2(x)\mathrm{d}x = \mathbb{E}\left\{\int_0^\infty \hat{f}^2_{m,n,\theta}(x)\mathrm{d}x\right\} - 2\mathbb{E}\left\{\int_0^\infty \hat{f}_{m,n,\theta}(x)f(x)\mathrm{d}x\right\}$, which depends on the unknown f. A common practice is to replace this last quantity by an estimator built from a data-driven procedure based on LSCV,

$$\text{LSCV}(m) = \int_0^\infty \hat{f}^2_{m,n,\theta}(x)\mathrm{d}x - \frac{2}{n}\sum_{i=1}^n \hat{f}^{(-i)}_{m,n,\theta}(X_i), \tag{10}$$

and then to infer the optimal degree m by considering $m_{\text{LSCV}} = \min_{m\in\mathbb{Z}^+} \text{LSCV}(m)$.

4 Simulation study

4.1 Comparing semiparametric and parametric fits under a EGPD model

We consider 100 samples of size $n = 300$ from the simple parametric EGPD model based on $G(u) = u^\kappa$ with $\kappa = 2, \sigma = 1$ and $\xi = 0.2$.

Figure 1 compares three methods to choose the polynomial degree m. The y-axis represents the different values of $m \in \{2, 4, ..., 100\}$, and the x-axis shows the boxplot of 100 optimal values obtained with the LSCV (see Equation (10)), the Integrated Absolute Error (IAE, $\int_0^\infty |\hat{f}(x) - f(x)|\mathrm{d}x$) and the Integrated Squared Error (ISE, $\int_0^\infty \left\{\hat{f}(x) - f(x)\right\}^2 \mathrm{d}x$) approaches, respectively. The LSCV method displays a higher variability. Still, the median value (solid black line in each boxplot) is around the same level, that is, $\bar{m}_{\text{LSCV}} = \bar{m}_{\text{ISE}} = 16$, and $\bar{m}_{\text{IAE}} = 14$. For the remaining analysis of this ideal case, we set $m = 16$.

Now, we want to compare the fit under two setups. In the first one, we make inference with the knowledge that the true model is based on G of the form $G(u) = G_\kappa(u) = u^\kappa$. In that setup, we denote the true model by f_κ. In the second setup, we approximate the nonparametric function G with Bernstein polynomials, see Equations (3,4). It leads to a semiparametric estimation. Intuitively, the inference should be better in the first setup. Still, the upper panels of Figure 2 indicate that our semiparametric EGPD fit (EGPD$_{m,n}$) based on

Figure 1: Boxplots of the optimal Bernstein polynomial degree m of the 100 simulated samples described in Section 4.1 obtained with the LSCV, ISE, IAE approaches

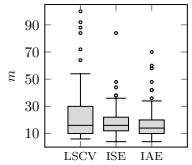
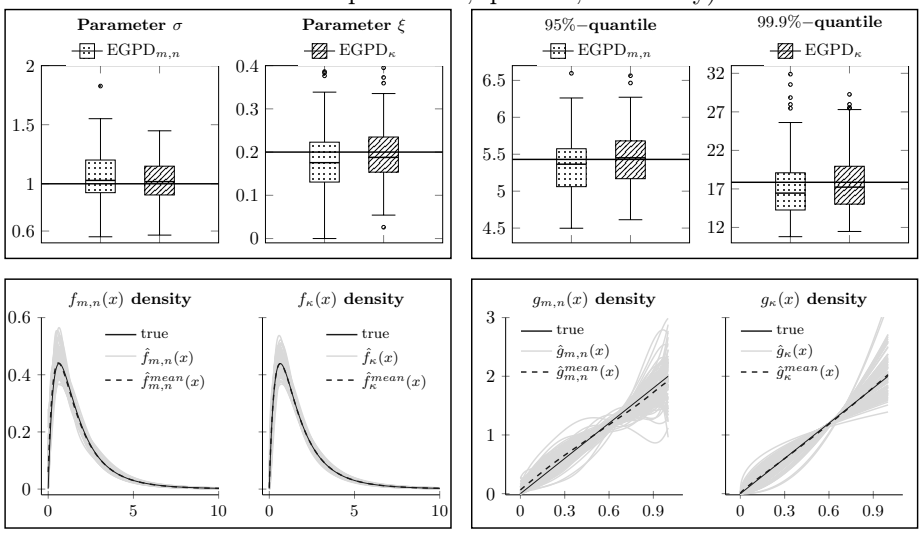


Figure 2: Comparing our semiparametric EGPD fit ($\text{EGPD}_{m,n}$) with a parametric EGPD fit when the true model is $G(u) = u^{\kappa}$ (EGPD_{κ}). Boxplots and densities are obtained from 100 samples of size $n = 300$ (the black solid lines indicate the true value of the parameter, quantile, or density)



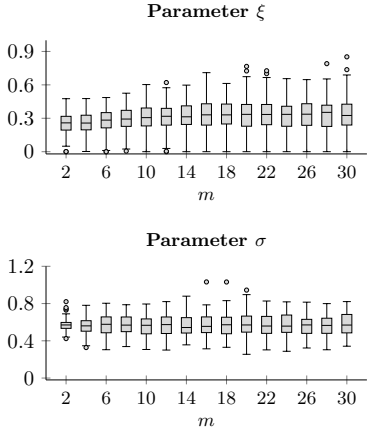
Bernstein polynomials (dotted boxplots) performs well for the inference of parameters (left) and quantiles (right) and it is competitive with the perfect model approach (dashed boxplots). Concerning the shape of the densities f and g, our Bernstein polynomials approach provides comparable estimates than the one obtained under the true model (lower panels of Figure 2).

4.2 Sensitivity study under a gamma-Fréchet mixture

Our second simulation study aims at assessing the quality of our approach when the data are drawn from a model outside of the EGPD class. More precisely, we consider the following gamma-Fréchet mixture

$$f_X(x) = pf_1(x) + (1-p)f_2(x)\,, \qquad \text{with } p \in [0,1], \tag{11}$$

Figure 3: Estimated (σ, ξ) parameters of the $\text{EGPD}_{m,n}$ model, considering 100 samples of size $n = 300$ from the gamma-Fréchet mixture defined by (11) with $p = 0.7$, $a = 1$, $b = 2.5$ and $\alpha = 2$.



and

$$\begin{cases} f_1(x) = x^{a-1} b^a \exp(-bx) / \Gamma(a), \ \forall\, x > 0, \\ f_2(x) = \alpha x^{-\alpha-1} \exp\left\{(-x)^{-\alpha}\right\}, \ \forall\, x > 0. \end{cases}$$

From EVT, we know that for any large threshold u and any $x \geq u$

$$\begin{aligned} \overline{F}(x) &= \mathbb{P}(X > u)\mathbb{P}(X > x | X > u), \\ &\approx (1 - q_u)\left(1 + \xi \tfrac{x-u}{\sigma_u}\right)^{-1/\xi}, \end{aligned}$$

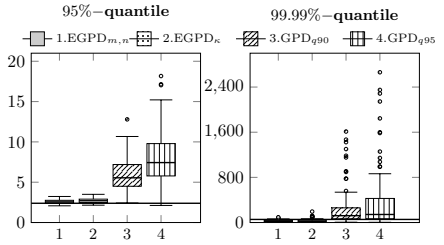
with q_u the quantile corresponding to the threshold u. As the upper tail of this mixture model is driven by the Fréchet component, we thus get, by identification, $\xi = \frac{1}{\alpha}$.

We draw 100 samples of size $n = 300$ from the model defined in Equation (11) with $p = 0.7$, $a = 1$, $b = 2.5$ and $\alpha = 2$ (other settings were also tested and they provide similar performance and interpretation). Our semiparametric EGPD model (denoted $\text{EGPD}_{m,n}$) is fitted to the 100 samples. To assess the influence of m on the estimation of $\boldsymbol{\theta} = (\sigma, \xi)$, these parameters are displayed for different degrees m in Figure 3. The choice of m does not appear to have a strong impact on the inference of σ and ξ in the case of the model defined in Equation (11).

As described in Section 3.4, in practice, the optimal polynomial degree m is obtained with the LSCV approach. This leads to $\bar{m}_{\text{LSCV}} = 8$, a degree that will be used for the remaining analysis of this case study. Note that the choice of m suggested in Babu *et al.* (2002), *i.e.*, $\bar{m}^n = \frac{n}{\log(n)}$, yields $\bar{m}^{n=300} = 53$, which clearly overfits in our framework.

With respect to extreme values, we compare our $\text{EGPD}_{m,n}$ model fit with the ones obtained from the simplest parametric EGPD with $G(u) = u^\kappa$ (denoted

Figure 4: Comparing the true values (in solid black) with the estimates of the quantiles obtained from 100 samples of size $n = 300$ and distributed from the gamma-Fréchet mixture defined by (11) with $p = 0.7$, $a = 1$, $b = 2.5$ and $\alpha = 2$.



EGPD$_\kappa$) and also from a classical GPD model. For the later, excesses are defined with respect to two classical threshold values: (a) the $90\%-$empirical quantile (GPD$_{q_{90}}$) and (b) the $95\%-$empirical quantile (GPD$_{q_{95}}$).

In this comparison exercise, our main goal is to assess the performance of each approach to infer very high quantiles, *i.e.,* their capacity to extrapolate beyond the largest observation. Figure 4 shows the boxplots of the estimated 95% and 99.99% quantiles. With respect to the true values (horizontal black lines), the EGPD$_{m,n}$ and EGPD$_\kappa$ models clearly outperform both classical GPD analysis, either based on excesses above the 90% quantile or the 95% quantile. The Root Mean Square Error (RMSE) shown in Table 1 confirms this result for even very large quantile estimates like 99.999%. We can also notice that our semi-parametric model EGPD$_{m,n}$ is superior to the simpler model EGPD$_\kappa$.

Overall, it is rather surprising that a semi-parametric model is better than a classical GPD approach as quantiles increase. This could be due to three reasons. First, our threshold choices (the 90% and 95% quantiles), although conventional in hydrology, were arbitrary and optimal threshold selection techniques used by EVT theoreticians may improve the GPD fit. Second, it seems that fitting adequately the bulk of the pdf via a Bernstein expansion, instead of being detrimental for extremes, brings an added value to the estimation of extreme quantiles. This remark may be specific to our gamma-Fréchet mixture, its pdf remains a very smooth function. Consequently, moderate extremes (*i.e.,* just below the threshold u) can still bring valuable information about the upper tail behavior, and reduces the variance of extreme quantile estimates. In contrast, the GPD approaches pay a heavy price in terms of variance because few points are above the threshold. Third, this table can only be interpreted in regards to this particular example and theoretical work is clearly needed to draw general conclusions.

Table 1: Root Mean Square Error (RMSE) between true and estimated quantiles. Each RMSE is computed by fitting four different models (columns) to 100 samples of size $n = 300$ drawn from the gamma-Fréchet mixture defined by (11) with $p = 0.7$, $a = 1$, $b = 2.5$ and $\alpha = 2$.

q	$\mathbf{EGPD}_{m,n}$	$\mathbf{EGPD}_{\kappa}$	$\mathbf{GP}_{q90}$	$\mathbf{GP}_{q95}$
90%	0.189	0.236	2.239	3.841
95%	0.324	0.447	3.949	6.528
99%	1.014	1.396	12.641	20.598
99.9%	6.522	7.977	67.549	114.569
99.99%	30.345	36.265	388.930	699.987
99.999%	118.002	143.920	2325.935	4496.131

Table 2: Daily rainfall data at Mont-Aigoual station in France (1976-2015): Seasonal σ and ξ estimates obtained with our semiparametric approach. The 95% confidence intervals (subscripts) are obtained from 100 non-parametric bootstrap replicates.

Season	$\hat{m}$	$\hat{\sigma}$	$\hat{\xi}$
Spring ($n = 494$)	24	$17.16_{[8.84,22.04]}$	$0.12_{[0.00,0.32]}$
Summer ($n = 352$)	18	$9.14_{[4.27,9.48]}$	$0.38_{[0.15,0.62]}$
Fall ($n = 514$)	18	$22.59_{[11.28,28.71]}$	$0.25_{[0.11,0.46]}$
Winter ($n = 538$)	22	$13.53_{[7.44,18.05]}$	$0.25_{[0.12,0.49]}$

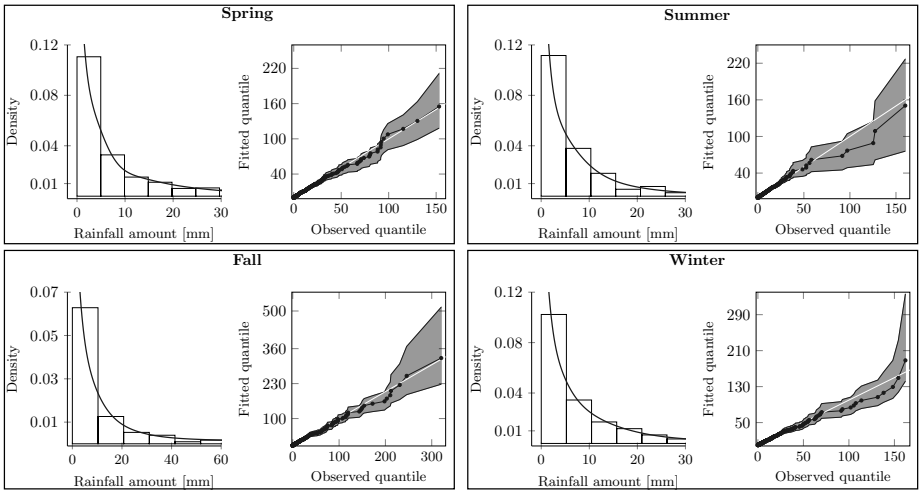
5 Application study - rainfall at Mont-Aigoual station

In this section, we apply our methodology on daily precipitation from 1976 to 2015 recorded at the Mont-Aigoual station in the south of France (e.g., see Carreau *et al.*, 2017, for a EVT analysis of this region). This location is known to produce very heavy rainfall and consequently, inference becomes a challenging task, especially the estimation of ξ.

A short exploratory analysis shows that there is a clear seasonal pattern, so each season is modeled separately: Spring (March-April-May), Summer (June-July-August), Fall (September-October-November), and Winter (December-January-February). To remove short-time temporal dependence, we only retain each third observation and also remove dry events (*i.e.*, zero precipitation values). After these steps, the sample sizes are per season: 494 in Spring, 352 in Summer, 514 in Fall and 538 in Winter.

The polynomial degree m is chosen with the LSCV approach presented in Section 3.4. For each season, LSCV is computed for each $m \in \{2, 4, \ldots, \frac{n}{\log(n)}\}$. Table 2 presents the estimated seasonal parameters $(\hat{\sigma}, \hat{\xi})$ obtained with our semiparametric model. While the estimated m does not vary with season, the two parameters are quite different, especially in Summer. We also observe the classical negative correlation between the two GPD parameters (see, e.g. Ribereau *et al.*, 2011).

Figure 5: Seasonal truncated histograms with the fitted densities and QQ-plots for the semiparametric $\text{EGPD}_{m,n}$ model for the daily rainfall data at Mont-Aigoual station (1976-2015). The gray bands represent the 95% confidence intervals.



To go further, we visually check the quality of the fit for each season. Empirical truncated histograms (ignore upper tail for a better visualisation) and estimated densities of f are illustrated in Figure 5. The quantile-quantile plots (QQ-plots) shade more light. In particular, the black curves are close to the diagonal, indicating that moderate and heavy rainfall are well captured by our semiparametric model, especially in Fall and Spring.

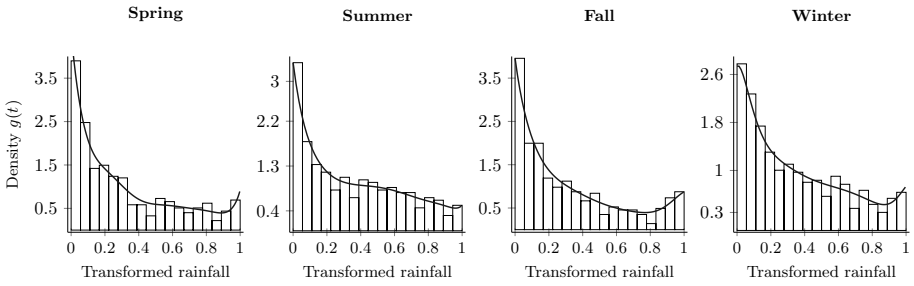
Using Equation (5), we recreate the hidden sample that should follow the pdf g. Figure 6 displays the impact of fitting our semiparametric model to this hidden sample. The solid black curves appear to provide a flexible fit, not only for the bulk of this hidden sample, but also for its upper tail, near one. This confirms our findings about the QQ-plot in Figure 5.

To summarize, the semiparametric model appears to fit reasonably well rainfall recorded at Mont-Aigoual station in France. The added flexibility of the semiparametric model is more visible in the QQ-plots and the fit of the hidden samples.

6 Conclusions and perspectives

This work addresses the statistical modeling of the entire range of precipitation amounts. The main benefit of our proposed approach is that low and large rainfall are in compliance with EVT and moderate precipitation are captured by a semiparametric model based on Bernstein polynomials. In other words, we have flexibility when we need flexibility (in the pdf bulk) and constraints when

Figure 6: Seasonal histograms with the fitted densities of the transformed daily rainfall at Mont-Aigoual station (1976-2015), see Equation (5), for the semi-parametric $\text{EGPD}_{m,n}$ method.



we need them (in the tails).

The performance of our semiparametric EGPD model has been evaluated with two simulation studies. For these examples, our inference method seems to accurately estimate even very large quantiles such as 99.999%.

From a computational point of view, selecting the appropriate Bernstein polynomial degree m by cross validation (LSCV) represents the most time consuming step. We therefore suggest the practitioner to explore first the range of admissible values for m on a coarse grid, and then to refine the search, by computing $\text{LSCV}(m)$ on a finer grid for m, around the optimum.

Concerning future work, our semiparametric EGPD model can be viewed as a "building block" for more complex statistical model (such as rainfall weather generators). In particular, the coupling with precipitation occurrences models could be very fruitful for assessment studies, like sensitivity of floods, erosion or crops models. On the same token, the modeling of the entire-range of precipitation amounts at multiple sites would be a welcome addition. More specifically, Evin *et al.* (2016) showed that a regional model can considerably improve the estimation of the GPD shape parameter.

Acknowledgement

The analysis presented in the case studies (simulation and rainfall data) of this work was performed in R Software (code provided upon request). P. Naveau would like to thank the LJK and IGE labs, as well as the Inria project/team AIRSEA, for hosting him for a month in Grenoble. This work has been partially supported by the Labex Persyval-Lab (ANR-11- LABX-0025-01) funded by the French program “Investissements d’avenir" through the exploratory project STAREX. program *Investissement d’avenir*. Part of this work was supported by the LEFE-INSU-Multirisk, ERC-A2C2 and EUPHEME projects. This work contributes to the CDP-Trajectories project, supported by the French National Research Agency in the framework of the “Investissements d’avenir" program (ANR-15-IDEX-02) The authors acknowledge Meteo France for the rainfall

dataset, that is available upon request. In addition, we wish to thank Valérie Monbet for her insightful suggestions and discussions.

References

Apipattanavis S, Podestã G, Rajagopalan B, Katz RW, 2007. A semiparametric multivariate and multisite weather generator. *Water Resources Research* **43**(11), w11401.

Babu G, Canty A, Chaubey Y, 2002. Application of Bernstein polynomials for smooth estimation of a distribution and density function. *Journal of Statistical Planning and Inference* **105**(2): 377–392.

Bernstein S, 1912. Démonstration du théorème de Weierstrass fondée sur le calcul des probabilités. *Communications de la Société mathématique de Kharkow* **13**(1): 1–2.

Bouezmarni T, Rolin J, 2007. Bernstein estimator for unbounded density function. *Journal of Nonparametric Statistics* **19**(3): 145–161.

Carreau J, Bengio Y, 2008. A hybrid Pareto model for asymmetric fat-tailed data: the univariate case. *Extremes* **12**(1): 53–76.

Carreau J, Naveau P, Neppel L, 2017. Partitioning into hazard subregions for regional peaks-over- threshold modeling of heavy precipitation. *Water Resources Research* **53**: 4407–4426.

Coles S, 2001. *An introduction to statistical modeling of extreme values*. Springer, 224 pp.

Davison AC, 1984. Modelling Excesses over High Thresholds, with an Application. In *Statistical Extremes and Applications*, Springer Netherlands, Dordrecht, 461–482.

Evin G, Blanchet J, Paquet E, Garavaglia F, Penot D, 2016. A regional model for extreme rainfall based on weather patterns subsampling. *Journal of Hydrology* **541**: 1185–1198.

Evin G, Favre AC, Hingray B, 2018. Stochastic generation of multi-site daily precipitation focusing on extreme events. *Hydrology and Earth System Sciences* **22**(1): 655–672.

Farouki R, 2012. The Bernstein polynomial basis: A centennial retrospective. *Computer Aided Geometric Design* **29**(6): 379–419.

Furrer EM, Katz RW, 2008. Improving the simulation of extreme precipitation events by stochastic weather generators. *Water Resources Research* **44**(12), w12439.

Garavaglia F, Gailhard J, Paquet E, Lang M, Garcon R, Bernardara P, 2010. Introducing a rainfall compound distribution model based on weather patterns sub-sampling. *Hydrology and Earth System Sciences* **14**(6): 951–964.

Ghosal S, 2001. Convergence rates for density estimation with Bernstein polynomials. *The Annals of Statistics* **29**(5): 1264–1280.

Grimshaw SD, 1993. Computing Maximum Likelihood Estimates for the Generalized Pareto Distribution. *Technometrics* **35**(2): 185–191.

Hegerl G, Zwiers F, 2011. Use of models in detection and attribution of climate change. *Wiley interdisciplinary reviews: climate change* **2(4)**: 570–591.

Ji Y, Wu C, Liu P, Wang J, Coombes K, 2005. Applications of beta-mixture models in bioinformatics. *Bioinformatics* **21**(9): 2118–2122.

Kakizawa Y, 2004. Bernstein polynomial probability density estimation. *Journal of Nonparametric Statistics* **16**(5): 709–729.

Katz R, 1977. Precipitation as a chain-dependent process. *Journal of Applied Meteorology* **16**(7): 671–676.

Katz R, Parlange M, Naveau P, 2002. Statistics of extremes in hydrology. *Advances in Water Resources* **25**(8): 1287–1304.

Leblanc A, 2010. A bias-reduced approach to density estimation using Bernstein polynomials. *Journal of Nonparametric Statistics* **22**(4): 459–475.

Leblanc A, 2012a. On estimating distribution functions using Bernstein polynomials. *Annals of the Institute of Statistical Mathematics* **64**(5): 919–943.

Leblanc A, 2012b. On the boundary properties of Bernstein polynomial estimators of density and distribution functions. *Journal of Statistical Planning and Inference* **142**(10): 2762–2778.

MacDonald A, Scarrott C, Lee D, Darlow B, Reale M, Russell G, 2011. A flexible extreme value mixture model. *Computational Statistics & Data Analysis* **55**: 2137–2157.

McLachlan G, Krishnan T, 2007. *The EM algorithm and extensions*, volume 382. John Wiley & Sons.

Nadarajah S, 2005. Extremes of daily rainfall in west central Florida. *Climatic Change* **69**(2-3): 325–342.

Naveau P, Huser R, Ribereau P, Hannart A, 2016. Modeling jointly low, moderate, and heavy rainfall intensities without a threshold selection. *Water Resources Research* **52**(4): 2753–2769.

Petrone S, 1999. Bayesian density estimation using Bernstein polynomials. *Canadian Journal of Statistics* **27**(1): 105–126.

Ribereau P, Naveau P, Guillou A, 2011. A note of caution when interpreting parameters of the distribution of excesses. *Advances in Water Resources* **34**(10): 1215 – 1221.

Richardson C, 1981. Stochastic simulation of daily precipitation, temperature, and solar radiation. *Water Resources Research* **17**(1): 182–190.

Stern R, Coe R, 1984. A model fitting analysis of daily rainfall data. *Journal of the Royal Statistical Society. Series A (General)* **147**(1): 1.

Vitale R, 1975. A Bernstein polynomial approach to density function estimation. *Statistical inference and related topics* **2**: 87–99.

Vrac M, Naveau P, 2007. Stochastic downscaling of precipitation: from dry events to heavy rainfalls. *Water Resources Research* **43**(7): 1–13.

Vrac M, Stein M, Hayhoe K, 2007. Statistical downscaling of precipitation through nonhomogeneous stochastic weather typing. *Climate Research* **34**(3): 169–184.

Wilks D, 1989. Conditioning stochastic daily precipitation models on total monthly precipitation. *Water Resources Research* **25**(6): 1429–1439.

Wilks D, 1999. Interannual variability and extreme-value characteristics of several stochastic daily precipitation models. *Agricultural and Forest Meteorology* **93**(3): 153–169.

Wilks D, 2011. *Statistical methods in the atmospheric sciences*. Academic Press, 676 pp.

Woolhiser D, Pegram G, 1979. Maximum likelihood estimation of Fourier coefficients to describe seasonal variations of parameters in stochastic daily precipitation models. *Journal of Applied Meteorology* **18**(1): 34–42.

Zucchini W, Adamson P, 1984. *The occurrence and severity of droughts in South Africa*. South Africa Water Research Commission.

A Appendix

[Proof of Lemma 1]

1. We write:

$$\frac{\hat{G}_{m,n}\left\{H_\xi(x/\sigma)\right\}}{x^s} = \frac{\hat{G}_{m,n}\left\{v\frac{H_\xi(v)}{v}\right\}}{\hat{G}_{m,n}(v)}\frac{\hat{G}_{m,n}(v)}{v^s}\sigma^{-s}$$

where $v = x/\sigma$. Note that $\lim_{v\to 0} \frac{H_\xi(v)}{v} = 1$. Thus, from the polynomial assumption on our model – $\hat{G}_{m,n}(t) = \sum_{k=0}^{m} G_n(\frac{k}{m}) b_{k,m}(t)$, $t \in [0,1]$ – we deduce:

$$\lim_{x\to 0} \frac{\hat{G}_{m,n}\{H_\xi(x/\sigma)\}}{x^s} = \lim_{v\to 0} \frac{\hat{G}_{m,n}\left\{v\frac{H_\xi(v)}{v}\right\}}{\hat{G}_{m,n}(v)} \frac{\hat{G}_{m,n}(v)}{v^s}\sigma^{-s} = \sigma^{-s}\lim_{v\to 0}\frac{\hat{G}_{m,n}(v)}{v^s}.$$

Then, from l'Hôpital's rule, and from Equation (2),

$$\lim_{v\to 0} \frac{\hat{G}_{m,n}(v)}{v^s} = \lim_{x\to 0} \frac{\hat{g}_{m,n}(v)}{sv^{s-1}}$$

is equivalent to $\frac{m\binom{m-1}{s-1}\omega_{s,m}v^{s-1}}{sv^{s-1}} = \frac{m\binom{m-1}{s-1}\omega_{s,m}}{s}$ with s the position of the first non-null weight in $\boldsymbol{\omega}$.

2. Let $u = \overline{H}_\xi(x/\sigma)$. We have

$$\lim_{x\to\infty} \frac{\overline{\hat{G}}_{m,n}\{H_\xi(x/\sigma)\}}{\overline{H}_\xi(x/\sigma)} = \lim_{u\to 0} \frac{\overline{\hat{G}}_{m,n}(1-u)}{u}.$$

Then, by applying l'Hôpital's rule, we get

$$\lim_{u\to 0} \frac{\overline{\hat{G}}_{m,n}(1-u)}{u} = \lim_{v\to 0} \hat{g}_{m,n}(1-u) = \hat{g}_{m,n}(1) = m\omega_{m,m}.$$

Therefore, to force the upper tail behavior in our model, we assume $\omega_{m,m} > 0$.

3. As the random variable Y can be written (in distribution) as $Y = \sigma H_\xi^{-1}\left\{G^{-1}(U)\right\}$, where U follows an uniform distribution on $[0,1]$, *i.e.,* $\mathbb{P}(U > w) = 1 - w$ for any $w \in [0,1]$, it follows that

$$\begin{aligned}
\mathbb{P}(Y > x+u | Y > u) &= \frac{\mathbb{P}(\sigma H_\xi^{-1}\left\{G^{-1}(U)\right\} > x+u)}{\mathbb{P}(\sigma H_\xi^{-1}\left\{G^{-1}(U)\right\} > u)}, \\
&= \frac{\mathbb{P}(U > G\left[H_\xi\left\{(x+u)/\sigma\right\}\right])}{\mathbb{P}(U > G\left\{H_\xi(u/\sigma)\right\})}, \\
&= \frac{1 - G\left[H_\xi\left\{(x+u)/\sigma\right\}\right]}{1 - G\left\{H_\xi(u/\sigma)\right\}}, \\
&= \frac{\overline{G}(1-w)}{\overline{G}(1-w^*)}, \text{ with } w = \overline{H}_\xi\left\{(x+u)/\sigma\right\} \text{ and } w^* = \overline{H}_\xi(u/\sigma), \\
&= \frac{\overline{G}(1-w)}{w}\frac{w^*}{\overline{G}(1-w^*)}\frac{w}{w^*}.
\end{aligned}$$

We assume that Y follows the classical EVT theory, *i.e.,* $\mathbb{P}(Y > x+u|Y > u)$, goes to $(1+\tilde{\xi}\frac{x}{\tilde{\sigma}})^{-1/\tilde{\xi}}$, as u gets large. Our constraint on G implies

that $\frac{\overline{G}(1-w)}{w}\frac{w^*}{\overline{G}(1-w^*)} \to 1$ for large u. So, the righthand side behaves as $\frac{w}{w^*} = (1+\xi\frac{x}{\sigma_u})^{-1/\xi}$ with $\sigma_u = \sigma + \xi u$. This implies that $\tilde{\xi} = \xi$ and $\tilde{\sigma} = \sigma_u$ for large u.