Estimation of uncertainties in Intensity Duration Frequency curves of extreme rainfall - a regional analysis.

Victor Mélèse [1] & Juliette Blanchet [2]

[1] *Univ. Grenoble Alpes, CNRS, IRD, Grenoble INP, IGE, F-38000 Grenoble, France*
victor.melese@univ-grenoble-alpes.fr
[2] *Univ. Grenoble Alpes, CNRS, IRD, Grenoble INP, IGE, F-38000 Grenoble, France*
juliette.blanchet@univ-grenoble-alpes.fr

Résumé. Nous proposons dans cette présentation, une étude régionale des incertitudes des courbes IDF calculées à partir de maxima ponctuels. Nous développons deux modèles de valeurs extrêmes généralisées valables de la résolution horaire à multi-journalière en utilisant une relation fractale entre les durées. Le premier est développé dans le cadre fréquentiste, le second dans le cadre bayésien. Dans le cadre fréquentiste les incertitudes sont obtenues i) par application du théorème de normalité asymptotique et ii) par bootstrap. Dans le cadre bayésien les incertitudes sont obtenues à partir de la densité a posteriori. Nous comparons ces deux cadres sur la même base de données qui englobe une région de 100000 km^2 située au Sud de la France. Cette région possède un régime de pluie très hétérogène ce qui nous permet de tirer des conclusions générales. Nous appliquons nos deux cadres sur 406 pluviomètres horaires dont les mesures commencent dans les années 80, pour une gamme de durées allant de 3 heures à 5 jours. Nous montrons que i) la densité a posteriori et la densité obtenue par bootstrap sont plus flexibles que la density gaussienne pour calculer les incertitudes et ii) la densité obtenue par bootstrap donne des intervalles de confiances irréalistes, en particulier pour les niveaux retour associés à de grande périodes de retour. Nous recommandons ainsi l'utilisation du cadre bayésien pour calculer les incertitudes.

Mots-clés. Environnement, climat, valeurs extrêmes, méthode bayésienne.

Abstract. We propose in this presentation a regional study of uncertainties in IDF curves derived from point-rainfall maxima. We develop two generalized extreme value models based on the simple scaling assumption, first in the frequentist framework and second in the Bayesian framework. Within the frequentist framework, uncertainties are obtained i) from the Gaussian density stemming from the asymptotic normality theorem of the maximum likelihood and ii) with a bootstrap procedure. Within the Bayesian framework, uncertainties are obtained from the posterior densities. We confront these two frameworks on the same database covering a large region of $100,000$ km^2 in southern France with contrasted rainfall regime, in order to draw conclusion not specific to the data. The two frameworks are applied to 406 hourly stations with data back to the 80's, accumulated in the range 3h-120h. We show that i) the posterior and the bootstrap densities are able to better adjust uncertainty estimation to the data than the Gaussian density, and ii) the bootstrap density give unreasonable confidence intervals, in particular for return levels associated to large return period. Therefore our recommendation goes towards the use of the Bayesian framework to compute uncertainty.

Keywords. Environment, climate, extreme values, Bayesian method.

1 Two frameworks of IDF relationships

1.1 Frequentist framework

1.1.1 Model

The frequentist framework is that considered in [1]. It relies on two assumptions. First, on the strict sense simple scaling assumption [2]

$$\mathrm{pr}(\boldsymbol{M}_D < x) = \mathrm{pr}\left\{\left(\frac{D}{D_{ref}}\right)^{-H} \boldsymbol{M}_{D_{ref}} < x\right\}, \tag{1}$$

where $\boldsymbol{M}_D$ is the random variable of annual maximum rainfall intensity for a duration D, $\boldsymbol{M}_{D_{ref}}$ is the random variable of annual maximum rainfall intensity for a duration of reference D_{ref} (D_{ref} = 3h in the application of Section 2), and H is a non-negative scalar called the scaling exponent.

The second assumption of our model is founded by extreme value theory [5] and asserts that annual maximum rainfall intensity at reference duration, $\boldsymbol{M}_{D_{ref}}$, follows a Generalized Extreme Value (GEV)

$$\mathrm{pr}(\boldsymbol{M}_{D_{ref}} < x) = \exp\left[-\left(1 + \xi\frac{x - \mu_{D_{ref}}}{\sigma_{D_{ref}}}\right)^{-\frac{1}{\xi}}\right], \tag{2}$$

provided $1 + \xi\frac{x-\mu_{D_{ref}}}{\sigma_{D_{ref}}} > 0$, where $\mu_{D_{ref}}$, $\sigma_{D_{ref}} > 0$, ξ are scalars, called respectively the location, scale and shape parameters. Case $\xi = 0$ corresponds to the Gumbel distribution

$$\mathrm{pr}(\boldsymbol{M}_{D_{ref}} < x) = \exp\left[-\exp\left(-\frac{x - \mu_{D_{ref}}}{\sigma_{D_{ref}}}\right)\right]. \tag{3}$$

(2) associated with (1) implies that annual maximum rainfall intensity $\boldsymbol{M}_D$ of any duration D follows a GEV distribution [1] and that the GEV parameters at duration D and D_{ref} are linked through $\mu_D = \left(\frac{D}{D_{ref}}\right)^{-H} \mu_{D_{ref}}$, $\sigma_D = \left(\frac{D}{D_{ref}}\right)^{-H} \sigma_{D_{ref}}$, while the shape parameter ξ does not depend on the time scale. As a consequence, the IDF relationships

relating the duration D, the return period T_R and the return level (i.e. the quantile of order $1 - 1/T_R$ of the corresponding GEV distribution) is given by

$$m_{D,T_R} = \left(\frac{D}{D_{ref}}\right)^{-H} \left\{ \mu_{D_{ref}} - \frac{\sigma_{D_{ref}}}{\xi} \left(1 - \left[-\log(1 - \frac{1}{T_R})\right]^{-\xi}\right)\right\}. \tag{4}$$

1.1.2 Inference

The unknown parameters to be estimated are $\theta = (\mu_{D_{ref}}, \sigma_{D_{ref}}, \xi, H)$. As in [1], θ is estimated by maximizing the likelihood under the assumptions that i) annual maxima are independent from one year to another, and ii) annual maxima of a given year at different durations are independent. This later assumption is likely to be miss-specified. For instance a 4h annual maximum is likely to be correlated with a 3h annual maximum. However incorporating dependence among many durations complicates much the modeling and its estimation, with little gain, if not loss, when only the marginal distributions are on interest [6]. The model log-likelihood is given by

$$\begin{aligned} l(\theta) = &\sum_{D \in \mathcal{D}} n(D) \log \left(\frac{D}{D_{ref}}\right)^{H} - \log(\sigma_{D_{ref}}) \sum_{D \in \mathcal{D}} n(D) - \\ &\frac{\xi + 1}{\xi} \sum_{D \in \mathcal{D}} \sum_{i=1}^{n} \log \left(1 + \xi \frac{\left(\frac{D}{D_{ref}}\right)^{H} m_{D,i} - \mu_{D_{ref}}}{\sigma_{D_{ref}}}\right) - \\ &\sum_{D \in \mathcal{D}} \sum_{i=1}^{n} \left[1 + \xi \frac{\left(\frac{D}{D_{ref}}\right)^{H} m_{D,i} - \mu_{D_{ref}}}{\sigma_{D_{ref}}}\right]^{-\frac{1}{\xi}}, \end{aligned} \tag{5}$$

where $n(D)$ is the number of observed years at duration D, $m_{D,i}$ is the annual maximum rainfall intensity at the duration D for year number i and $\mathcal{D}$ is the set of considered durations. There is no analytical form for the maximum of l but maximization can be obtained numerically (e.g. quasi Newton method).

1.1.3 Uncertainty computation

We propose two ways to computing uncertainty in the frequentist framework. The first one relies on the asymptotic normality of the maximum likelihood estimator, but using the correction described in [4] to account for the fact that the likelihood (5) ignores dependence among maxima of the same year. The second one is based on bootstrap resampling.

1.2 Bayesian framework

1.2.1 Model

As in the frequentist framework, the Bayesian framework relies on the strict sense simple scaling hypothesis combined with the GEV distribution. However in this case, the model parameters $\boldsymbol{\theta} = (\boldsymbol{\mu}_{D_{ref}}, \boldsymbol{\sigma}_{D_{ref}}, \boldsymbol{\xi}, \boldsymbol{H})$ are random variables. Thus the two above hypothesis, as all the equations derived in Section 1.1.1, still apply but conditionally on $\boldsymbol{\theta}$ equals to some $\theta = (\mu_{D_{ref}}, \sigma_{D_{ref}}, \xi, H)$. Thus, conditional on $\boldsymbol{\theta} = \theta$, the annual maximum rainfall intensity $\boldsymbol{M}_D$ of any duration D, follows a GEV distribution, i.e.

$$\text{pr}(\boldsymbol{M}_D < x | \boldsymbol{\theta} = \theta) = \exp\left[-\left(1 + \xi \frac{x - \mu_D}{\sigma_D}\right)^{-\frac{1}{\xi}} \right], \tag{6}$$

where $\mu_D = \left(\frac{D}{D_{ref}}\right)^{-H} \mu_{D_{ref}}$ and $\sigma_D = \left(\frac{D}{D_{ref}}\right)^{-H} \sigma_{D_{ref}}$.

1.2.2 Inference

In the Bayesian framework, interest is in estimating the density of the parameters knowing the data, called the posterior density. The well known Bayes formula states that

$$f(\theta|m) = \frac{f(m|\theta)f(\theta)}{\int_\theta f(m|\theta)f(\theta)d\theta}, \tag{7}$$

where $f(\theta)$ is the prior density, and $f(m|\theta)$ is the density associated to the data under (6).

In our case, as often in Bayesian analysis, there is no analytical form for the posterior density (7) due to the presence of an integral in the normalizing constant. This problem can be overcome by using simulation based techniques such as Markov chain Monte Carlo (MCMC), which provides a way of simulating from complex distributions, such as $f(\theta|m)$, by simulating from Markov chains which have the target distributions as their stationary distributions. We use the Metropolis algorithm to sample $f(\theta|m)$. Uncertainty are obtained from the posterior density.

2 Results

2.1 Uncertainty estimation: the example of Montpellier

Fig. 1 illustrates that the posterior and bootstrap densities are able to better adjust to the data by being able to produce asymmetric densities with several modes. This produces asymmetry in return levels with a heavier right tails for the bootstrap and posterior densities than for the Gaussian density, whereas the left tails of the posterior and Gaussian

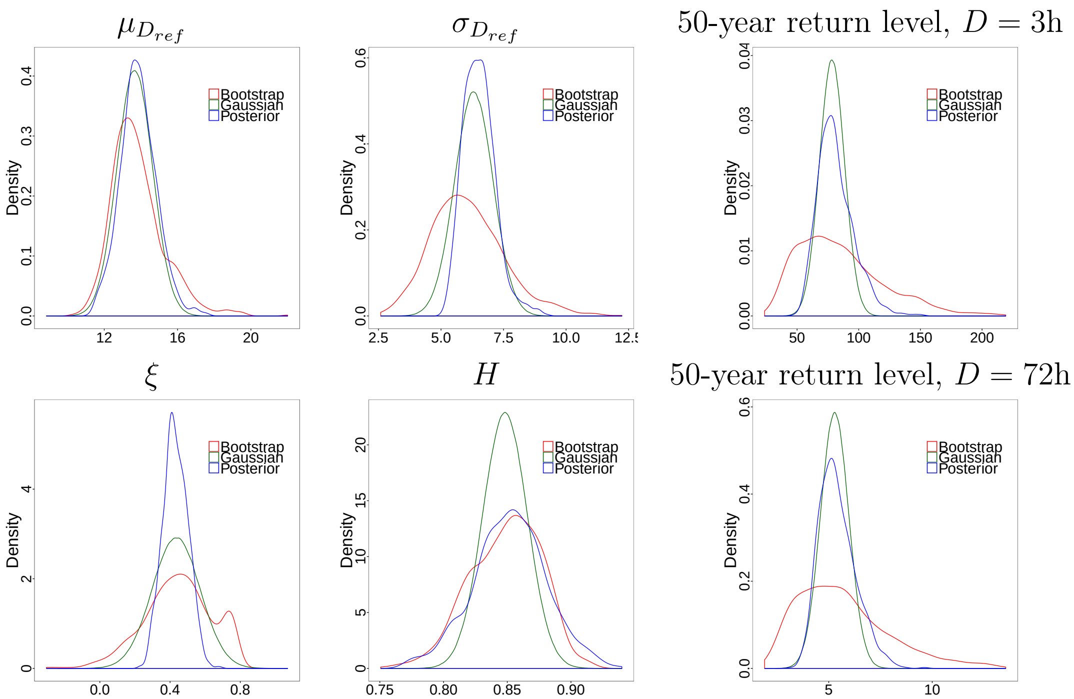

Figure 1: Density estimates of the model parameters and the 50-year return levels at 3h and 72h durations, for Montpellier station. Frequentist densities are obtained with the theorem of asymptotic normality (green) and the bootstrap resampling method (red). Bayesian densities are the posterior densities (blue).

densities are similar. Therefore the bootstrap and Bayesian methods are able to tell there is a greater likelihood for the 50-year return level to be above the estimated value than below it, which is not possible when considering symmetric Gaussian densities.

2.2 Regional study of uncertainty

Fig. 2 shows the skewness of the bootstrap and Bayesian densities for the all region. For sake of readability, we represent the Kernel densities of the skewness and restrict the x-axis to comprise 95% of the values. For the GEV parameters, most skewness of the posterior densities are positive, meaning heavier right tails. This also applies to the bootstrap densities but to a lesser extent to ξ. For the scaling parameter, both left and right heavy tails are found with both methods. For the return levels, mainly positive skewness are found. For the great majority of the stations, there is a greater likelihood for the 50-year return level to be above its estimated value than below it. This piece of information is of great importance for risk management and is missing when considering symmetric Gaussian densities stemming from the asymptotic normality theorem. The

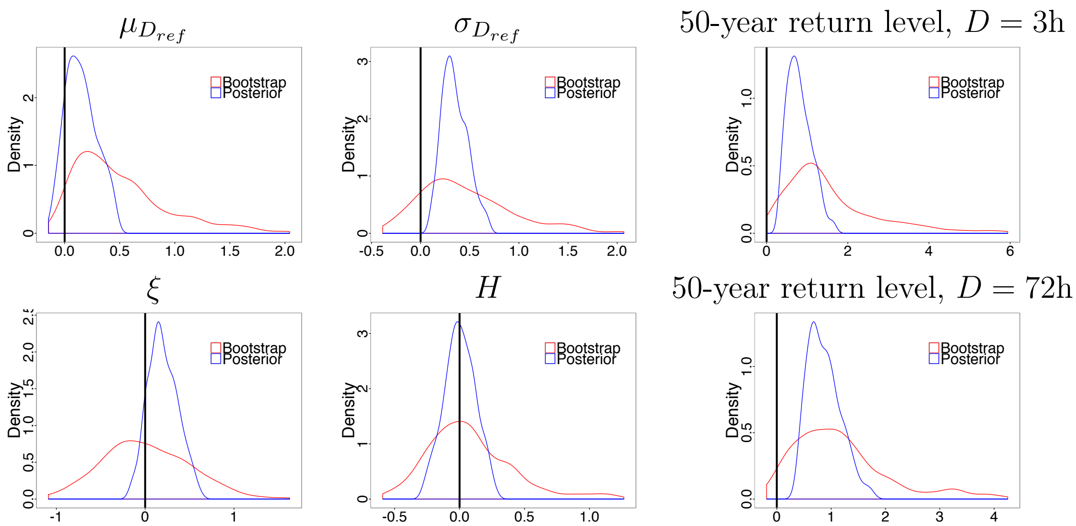

Figure 2: Regional skewness of the bootstrap (red) and posterior densities (blue) of the model parameters and the 50-year return levels at 3h and 72h durations. The black vertical line at 0 is the value for any symmetric density, such as the Gaussian density.

bootstrap skewness of all variables often largely exceed the Bayesian values. We can postulate that the bootstrap method tends to give too heavy right-tailed densities and are not recommended for the computation of uncertainty.

Fig. 3 compares the lower and upper bounds of the 95% confidence interval of the Gaussian and posterior densities for 3h-duration. It shows that the lower bounds are usually similar in both cases whereas the upper bounds of the posterior density are always greater. This corroborates the results found for the station of Montpellier in Section 2.1: the Bayesian framework allows to obtain assymmetric confidence bands extending further to large values. We conclude from Fig. 3 that the Gaussian density tends to underestimate uncertainty across the whole region.

Bibliographie

[1] Blanchet, J. and Ceresetti, D. and Molinié, G. and Creutin, J-D. (2016), A regional GEV scale-invariant framework for Intensity-Duration-Frequency analysis, *Journal of Hydrology*.
[2] Gupta, Vijay K. and Waymire, Ed. (1990), Multiscaling properties of spatial rainfall and river flow distributions, *Journal of Geophysical Research: Atmospheres*.
[3] Huard, D. and Mailhot, A. and Duchesne, S. (2003), Bayesian estimation of intensity–duration–frequency curves and of the return period associated to a given rainfall event, *Stochastic Environmental Research and Risk Assessment*.

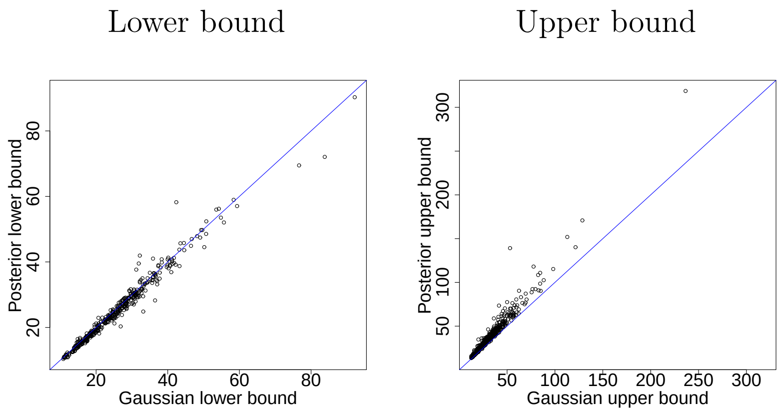

Figure 3: Comparison of the 95% confidence intervals of Gaussian and the posterior density for the 50-year return level at 3h-duration.

[4] Davison, A. C. (2008). Statistical Models. *Cambridge University Press, 1 edition*
[5] Coles, S. and Bawa, J. and Trenner, L. and Dorazio, P. (2001). An introduction to statistical modeling of extreme values. *Springer*
[6] Sebille, Q. and Fougères, A.-L. and Mercadier, C. (2016). A comparison of spatial extreme value models. Application to precipitation data. *preprint*