We warmly thank the Reviewer for his valuable comments. We provide below a detailed point-to-point reply to these comments. The proposed changes for the next version of the article are indicated in red.

GENERAL COMMENTS

This paper deals with the statistical modeling of the distribution of rainfall amounts within a region, especially focusing on extremes. The approach is computationally intensive, with the parameters of the rainfall distribution at individual sites being mapped across the region using (e.g.) Kriging or splines. Cross validation is applied to evaluate the performance of the candidate models (e.g., form of distribution and spatial interpolation technique). Challenges include the attempt to model the entire range of rainfall amounts (i.e., from near zero to the most extreme) with a single distribution.

It is claimed that the proposed approach to rainfall modeling is "general and could be applied to any region of the world" (p. 24, line 29). Yet some aspects of the approach seem tailored to the application to a specific region in France. In particular, seasonality is treated by dividing the year into two seasons, one in which extremes typically occur. Plus the model is fitted conditional on one of three possible weather patterns (WPs), based on the spatial correlation of rainfall for the region. Although the number of seasons and WPs could certainly be varied to model rainfall for other types of climate, it is not clear that the constraint of being limited to a quite small number of seasons and WPs could always permit an adequate fit.

⇒ We think there is a misunderstanding here. We fully agree that the seasonal and weather pattern modeling may not be relevant in other region of the world, but this is precisely what we claim p. 24 lines 29-33: "Although our procedure of selection is general and could be applied to any region of the world -and possibly to other variables (temperature etc)-, we stress that our conclusions are in themselves not universal. In particular, other marginal distributions may be more suitable than the Gamma in other regions of the world showing less or more heavy tails. Although the mixture of distributions over weather patterns has revealed efficient in other countries (e.g. in Norway and Canada, Brigode et al., 2014; Blanchet et al., 2015), it might be less relevant in, e.g., monsoon climate regions where the consideration of seasons seems essential and might be enough." So the term "general" refers to the selection procedure (i.e. to the cross-validation framework), not to the modeling. We claim that the proposed framework, as being based on objective criteria, is general and may be used to select among any distribution. Here we use mixtures of Gamma over seasons and weather patterns, but any other distribution may be considered. The proposed cross-validation criteria are general and independent on the choice of distributions.

In order to clarify this in the article, we propose to add p. 2 line 33 the following sentences: "... in very short distance. Following previous studies in the region (Evin et al., 2016; Garavaglia et al., 2010, 2011; Gottardi et al., 2012), the compared marginal distributions involve seasonal and weather pattern subsampling, considering different models for the subclass-dependent distributions. However the proposed cross-validation framework is general, as involving objective criteria, and could likewise be used to select among any other distribution.".

Another potential limitation concerns the performance of the different forms of distributions fitted to rainfall amounts, particularly for extremes. Conclusions are drawn about "heavy tails" that could benefit by relying more on extreme value theory. The restriction to a single distribution may have distorted the performance for extremes, with some of the conclusions conflicting with results in the literature when only extremes are modeled.

⇒ By "the restriction to a single distribution", we guess you refer to using a single distribution for modeling the whole range of rainfall. As already stated p. 2 lines 18-20, another possibility would indeed be to consider hybrid distributions built by mixing a first distribution representing the bulk of the distribution and a second one focusing on the upper tail and relying on Extreme Value Theory (see the review of Scarrot and MacDonald, 2012, for example). For example Frigessi et al. (2002) consider a mixture of Weibull and Generalized Pareto distributions modeling respectively the bulk and the heavy tail of the distribution. Frigessi's model presents the advantage of removing the delicate choice of a predetermined threshold. However, as noted by Naveau et al. (2016), it also has many parameters (6) which are difficult to estimate. We don't think that applying such a model in the context of WP and season subsampling is relevant. This is why we preferred considering the extended Generalized Pareto model of Naveau et al. (2016) which i) allows to model both the bulk and the tail of the distribution, ii) is in compliance with Extreme Value Theory, iii) is much more parsimonious (3 parameters). Given the lack of robustness of the extended Generalized Pareto distribution (see the SPAN100 scores of Figure 6), we firmly believe that Frigessi's model (or equivalent) would lack robustness even more. Regarding the fact that the selected model is not in compliance with Extreme Value Theory, let us point out that the mixture of Gamma distribution looks actually "pretty much" like the extended Generalized Pareto distribution (compare Figure 7 to Figure 5) but it is much more robust. Finally let us remind that Extreme Value Theory is an asymptotic theory, so it applies to peaks over infinite threshold, which is obviously never the case in practice. However for long-enough data one can usually reasonably assume convergence to the asymptotic case and base analysis of extremes on Extreme Value Theory. So the fact that the distribution founded by Extreme Value Theory is not selected in our case might be an indication that the available data are not long-enough to consider the asymptotic theory to hold.

For these reasons, I recommend that the manuscript be accepted for publication subject to revision.

SPECIFIC COMMENTS

(1) Generality of proposed approach

It seems like a crude approximation to consider only two seasons and assume stationarity within a given season. More realistic approaches include allowing the parameters of the rainfall distribution to gradually *change depending on the time of year. Some regions of the world even have more than one wet season, indicating a limitation of the proposed approach.*

⇒ Actually we assume stationarity within a given season **and** weather type. Figure 4 of the paper shows there is a clear gain in considering the weather patterns (WPs) to complement the information brought by subsampling into seasons. This said, we fully agree that considering two fixed seasons might not be relevant in other region of the world. This is stated in the discussion section, see p. 24 line 31 to p. 25 line 1: "Although the mixture of distributions over weather patterns has revealed efficient in other countries (e.g. in Norway and Canada, Brigode et al., 2014; Blanchet et al., 2015), it might be less relevant in, e.g., monsoon climate regions where the consideration of seasons seems essential and might be enough." However the use of fixed seasons and WPs has already been extensively studied and validated in southern France in other studies (Evin et al., 2016; Garavaglia et al., 2010, 2011; Gottardi et al., 2012). Considering rainfall distributions that gradually change depending on the time of year would be another possibility but it would be tricky in our case. Figure 1 shows the monthly averages of daily nonzero rainfall in the region. It clearly shows the occurrence of a wetter season spanning the months of September, October and November. These three months compose the season-at-risk considered in this paper. A second maximum, although much lower, is found in April and May. Modeling the monthly fluctuations of Figure 1 as a parametric function of time (for example using cos and sin terms) doesn't seem easy to us. Fitting one distribution per month would be another possibility but it would very likely lack robustness by requiring estimating 12 distributions. Considering WPs is an alternative way of modeling the monthly fluctuations through the monthly occurrence of the WPs (see Figure 1). It shows the advantage of being physically-based since the original WPs of Garavaglia et al. (2010) are constructed by clustering synoptic circulations (geopotential fields). Considering WPs allows in particular accounting for the occurrence of two wetter periods -firstly in automn (season-at-risk) and secondary in spring (in the season-not-at-risk)-, which correspond to larger probabilities of occurrence of WP1, which is the wettest WP for both seasons (see Figure 10 of the article for the case of the season-at-risk). This shows the advantage of being relatively parcimonious (it requires estimating 6 distributions in the WP/season subsampling case).

To make this clearer, we propose to replace Table 4 by a new Figure 4 corresponding to Figure 1 of this response and to replace p. 14 lines 14-16 by: "The occurrence statistics of the three WPs for the period 1948-2013 are presented in Figure 4. The yearly occurrence of the three WPs is roughly similar (27% for WP1, 36% for WP2, 37% for WP3). However the WPs show very different seasonality. In particular WP1 is more frequent in spring and autumn, which correspond to wetter periods, particularly in autumn (see the monthly averages of nonzero rainfall in Figure 4). WP3 is more frequent in summer, which is the dryest season, while WP2 features almost a reversed seasonality compared to WP3. This shows that, although being based on the spatial dependence, the WPs are linked to the seasonality of rainfall in the region. We also propose to clarify that the WPs are physically-based by adding p. 14 line 11: "... classification described in

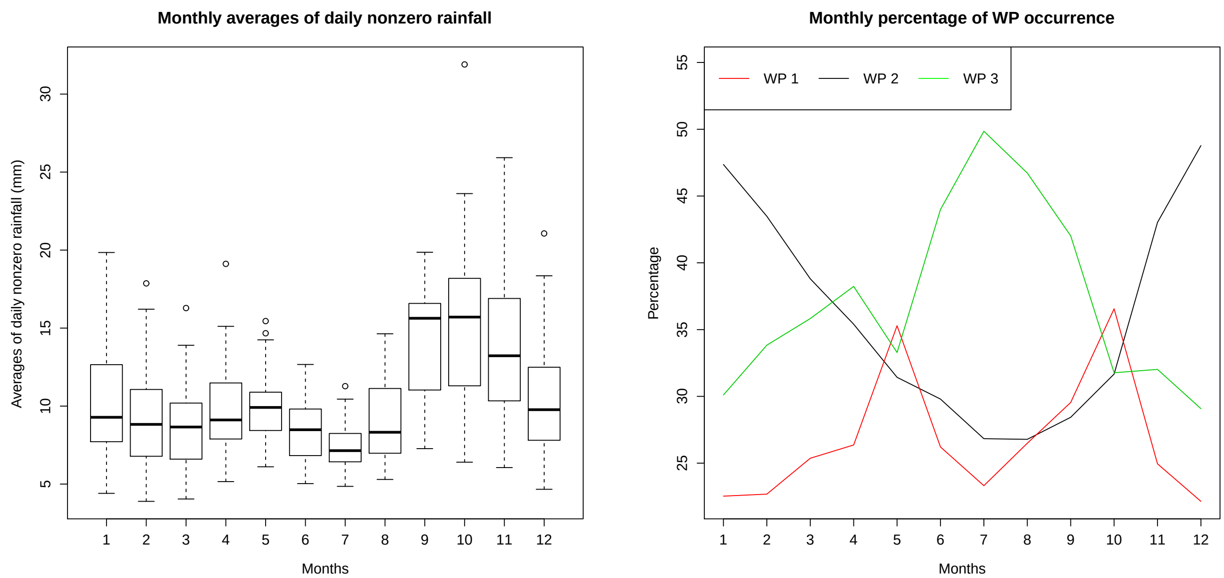

Figure 1: Left: Boxplot of the monthly averages of daily nonzero rainfall. Each boxplot contains 42 points (one point per station). Right: Monthly percentage of occurrence of the three WPs.

Garavaglia et al. (2010), which is obtained by clustering synoptic situations (geopotential heights) for France and surrounding areas into eight classes".

Conditioning on a few WPs based on the degree of spatial correlation of rainfall is an intriguing and not very common approach. Alternatives in the literature have included either introducing a hidden state variable (likewise assuming only a few possible states), which would require much more involved calculations, or including an observed covariate (such as an index of atmospheric circulation), which would require its identification for a given region but could assume effectively infinitely many possible states. Other than convenience, the advantages of the proposed approach are not clear.

⇒ This is perfectly right: as stated p. 14 lines 12-13, the three considered WPs are obtained by grouping the eight WPs of Garavaglia et al. (2010) into three classes based on their spatial correlation. However let us recall that the spatial correlation is only a secondary ingredient in the construction of the WPs. The first ingredient is the similarity in synoptic situations since the eight WPs of Garavaglia et al. (2010) are obtained by clustering geopotential fields. Thus the used WPs do already include atmospheric information. This said, we understand it might seem intriguing to consider the spatial correlation for the subgouping whereas this article deals with marginal distribution. Actually this choice is guided by external constraints that we deliberately omitted to mention to not confuse the reader. Indeed, this study is part of a bigger project aiming to build a stochastic daily rainfall generator in the region. The first step of this project was to select an appropriate marginal distribution that could reasonably fit the whole range of rainfall values anywhere within the region. This has led us to develop the cross-validation framework of this article. However having in mind that the final goal will be to generate rainfall fields and that this generation will be based on WPs, we decided to reduce the number of parameters of the rainfall generator by grouping some of the WPs based on the spatial correlation of rainfall during these WPs. The maps of Figure 10 of the paper lead us conclude that this grouping, although based on the spatial correlation, makes actually good sense for the marginal distribution. There may be other possible groupings but we repeat that the goal of this paper is *not* to find the best marginal model but to propose an objective cross-validation framework for selecting among several marginal distributions.

(2) Using extreme value theory to interpret results

It is concluded that a mixture of gamma distributions provides the best fit, especially for extreme high precipitation amounts. Yet a gamma distribution has a light tail, well known to not be heavy enough for precipitation extremes. Still it is argued that allowing the gamma distribution to vary depending on the season and on the WP induces a heavier tail (Figure 5).

If this claim were correct for seasonality, then it would appear that the apparent heavy tail is at least partly an artifact of ignoring seasonality. Yet there is some evidence in the literature (e.g., by explicitly allowing some of the parameters in an extreme value analysis to vary within the year) that this is not necessarily the case.

Concerning conditioning on WPs, it is well known that a mixture of gamma distributions can induce a heavier tail than a single gamma. Yet I wonder whether a mixture involving only a few gamma distributions (i.e., only three for the wet season) is sufficient to produce a truly heavy tail (in the sense of extreme value theory).

So it may be informative to examine how well the gamma distribution (and the mixture of three gamma distributions depending on the WP) fits precipitation amounts in the wet season alone. As it stands, I worry that the results for extremes may have been distorted by the constraint of fitting a single distribution to all rainfall amounts.

⇒ This is a very good point. Let consider the marginal distribution of daily rainfall in the season-at-risk. Following Equation (3) of the article, the marginal distribution in season s is given by

$$\mathrm{pr}_s(R \leq r) = p_s^0 + \sum_{k=1}^{K} p_{s,k}(1 - p_{s,k}^0) G_{s,k}(r), \tag{1}$$

where $p_s^0 = \sum_{k=1}^{K} p_{s,k} p_{s,k}^0$ is the probability of any day to be dry in the season s. Nonzero precipitation amounts defined by (1) have CDF:

$$G_s(r) = \sum_{k=1}^{K} p'_{s,k} G_{s,k}(r), \tag{2}$$

where $p'_{s,k} = p_{s,k}(1 - p^0_{s,k})/(1 - p^0_s)$. We show in Figures 2 and 3 the cross-validation scores associated to (2) when s is the season-at-risk. Comparing Figure 2 below to Figure 4 of the article reveals the same conclusion, namely also for the season-at-risk there is a strong gain in considering WP subsampling, which applies both for the bulk (NRMSE) and the tail (FF and N_5) of the distribution, despite a slight loss in robustness ($SPAN_{100}$). Comparing Figure 3 below to Figure 6 of the article reveals also that, for the season-at-risk as well, the two best distributions are the extended Generalized Exponential (eexp) and the Gamma (gamma) distributions. We propose to add in the article:

- p. 6 line 7 "... $S \times K$ Gamma distributions. Analogously, the CDF of nonzero precipitation amounts in a given season s writes

$$G_s(r) = \mathrm{pr}(R \leq r | R > 0, \text{season} = s) = \sum_{k=1}^{K} p''_{s,k} G_{s,k}(r), \tag{5}$$

where $p''_{s,k} = p_{s,k}(1 - p^0_{s,k})/(1 - \sum_{k=1}^{K} p_{s,k} p^0_{s,k})$."

- p. 20 line 10 "... similarly for Antraigues station. Note that exactly the same conclusions hold when focusing on the season-at-risk rather than considering the whole year, i.e. when computing the cross-validation scores for the estimated seasonal distribution G_s in (5) rather than for the year-round distribution G in (4)."

However, for the sake on concision, we propose not to show any of the Figure 2 or 3 below.

(3) Assumption of temporal independence

It is effectively assumed that the rainfall amounts at an individual site, especially extreme high values, are temporally independent (e.g., second displayed equation on p. 8 and p. 9, line 7). But this assumption never appears to be explicitly stated or verified. There is some evidence in the literature of "clustering" at high levels for time series of daily rainfall amounts at individual sites. Cross validation, depending on how it is implemented, would not properly account for the effects of such temporal dependence.

⇒ Actually temporal dependence is weak in the region. If we define the wet periods of a given station as the number of consecutive days with nonzero daily rainfall, we find that 43% of the wet periods over the whole region have length 1, 25% have length 2 and 13% have length 3. To focus more on high levels, we computed the extremal index at each station using the method of Ferro and Segers (2003) (R package texmex), considering exceedances above the 90%- and 95%-quantiles. The regional average of the extremal indices amounts 0.65 for the 90%-quantile and 0.68 for the 95%-quantile. This means that the average cluster length of rainfall exceeding high levels is about $1/0.65 \approx 1.5$, which is close to independence (in which case the extremal index is 1). Therefore we think that the hypothesis of temporal independence is defensible for our data. However,

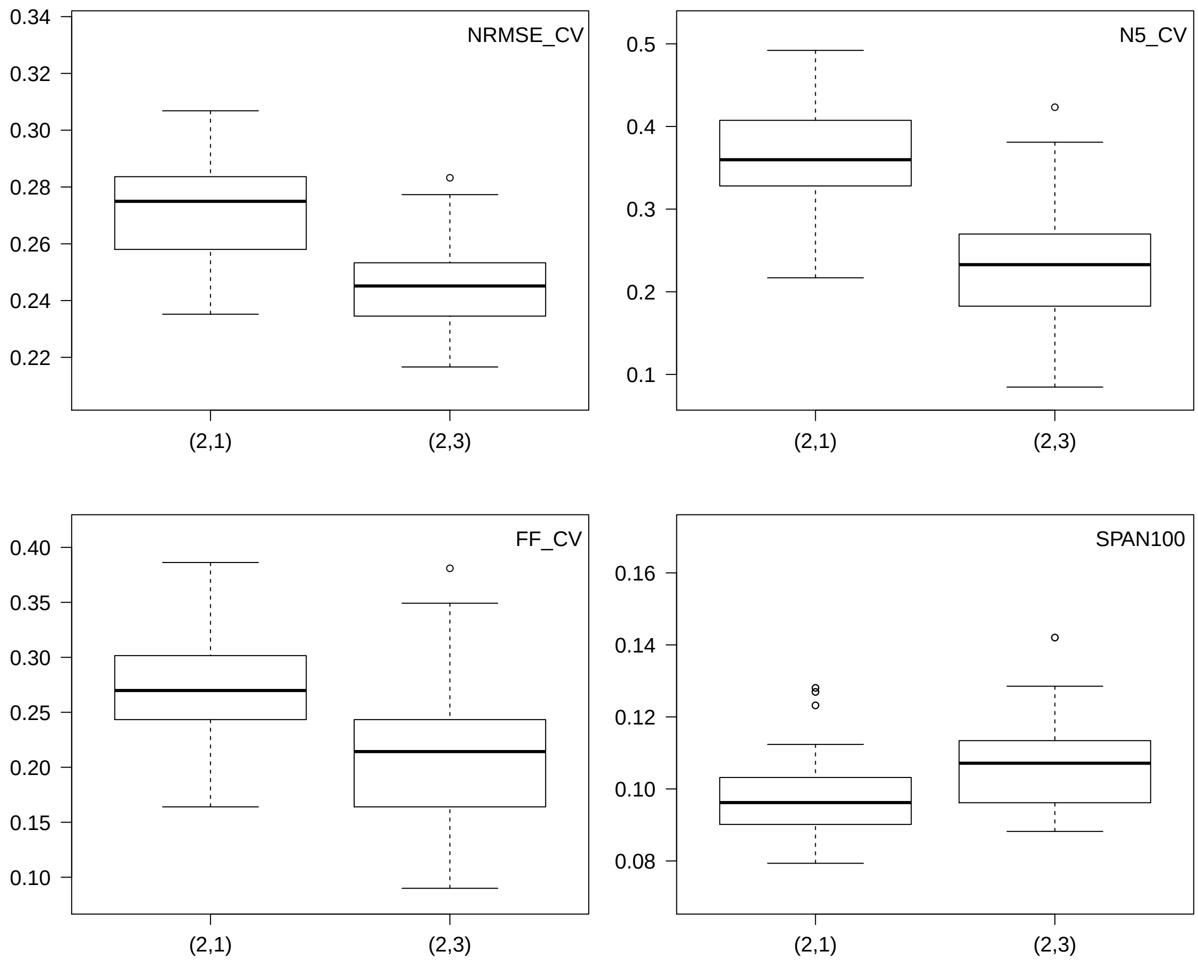

Figure 2: Scores of cross-validation in the season-at-risk when $G_{s,k}$ are Gamma distributions and the number of WT varies: $K \in \{1,3\}$. The values of (S,K) are indicated in the x-labels. Each boxplot contains 100 points.

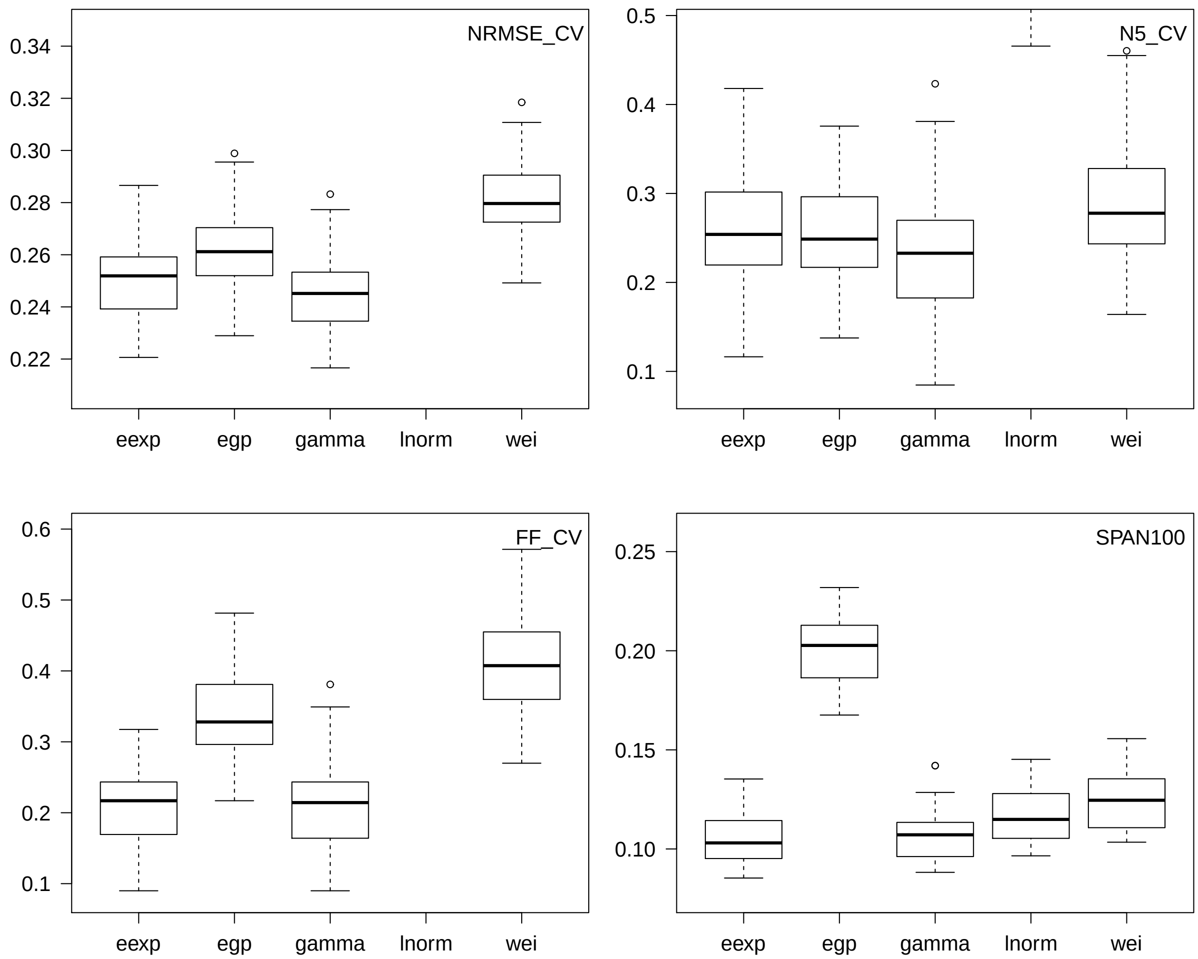

Figure 3: Scores of cross-validation in the season-at-risk when $G_{s,k}$ is either the extended exponential (eexp), extended Generalized Pareto (egp), Gamma (gamma), lognormal (lnorm) or Weibull (wei) distribution, with $S = 2$ and $K = 3$. Each boxplot contains 100 points. The boxplots of reliability scores in the lognormal case are missing because they lie far above the upper range of depicted values.

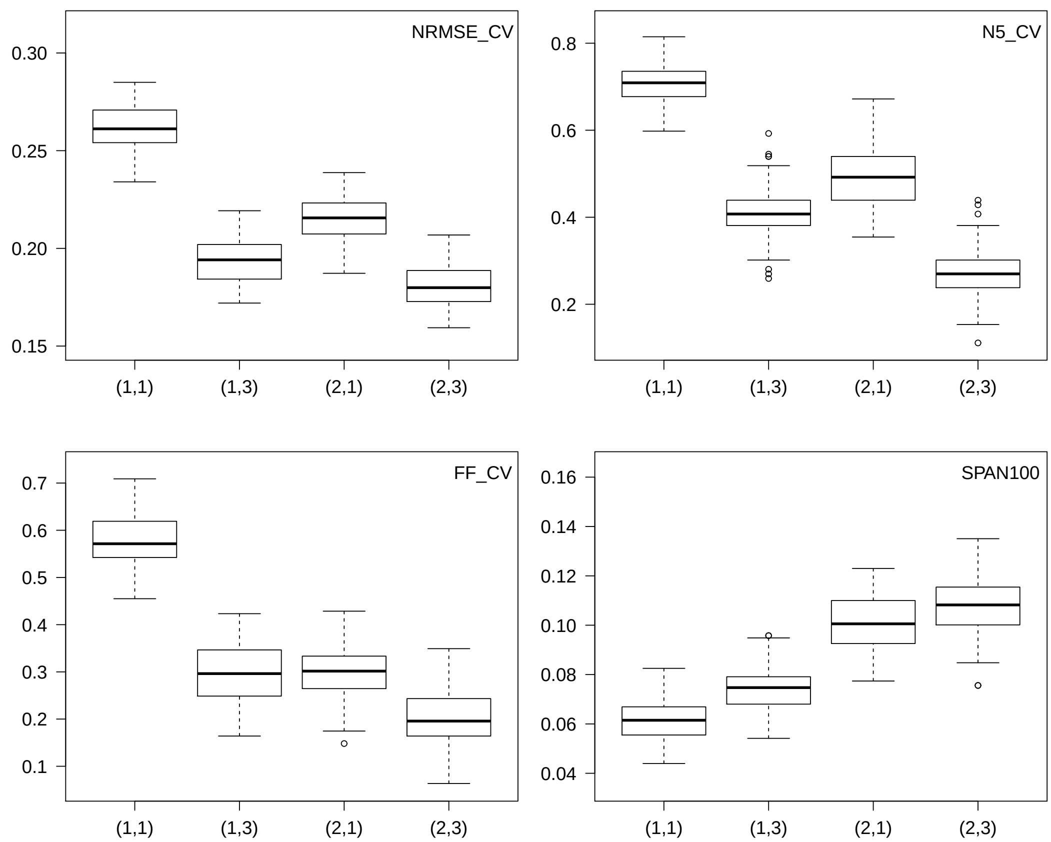

Figure 4: Scores of cross-validation obtained when split sampling blocks of five consecutive days, when $G_{s,k}$ are Gamma distributions and the number of seasons and WT varies: $S \in \{1,2\}$ and $K \in \{1,3\}$. The values of (S,K) are indicated in the x-labels. Each boxplot contains 100 points.

in order not to be biased in the cross-validation procedure by possible weak temporal dependence, we re-made all the estimation but split sampling for the cross-validation blocks of five consecutive days rather than individual days (still imposing half the data to be in the calibration sample and the remaining half to be in the validation sample). Results are shown in Figures 4 to 9, corresponding respectively to Figures 4 to 8 of the article. All results are actually almost similar to those of the article. Therefore the same conclusions hold, namely that i) there is some gain in considering subsampling into seasons and WPs (Figure 4 below), ii) the Gamma and extended Exponential models give overall the best scores of cross-validation, iii) the bivariate thin plate spline with drift in smoothed altitude (tps2Z) is the best interpolation method.

We propose to replace line 1 of p. 15 by: "1. We divide the days of 1948-2013 into two subsamples of equal size, denoted $C^{(1)}$ and $C^{(2)}$. Given the weak temporal dependence of rainfall in the region (80% of the wet periods have length lower than 3), division is made by randomly choosing blocks of five consecutive days to compose $C^{(1)}$, the remaining blocks composing $C^{(2)}$. Figures 4 to 8 of the article will be replaced by the corresponding figures.

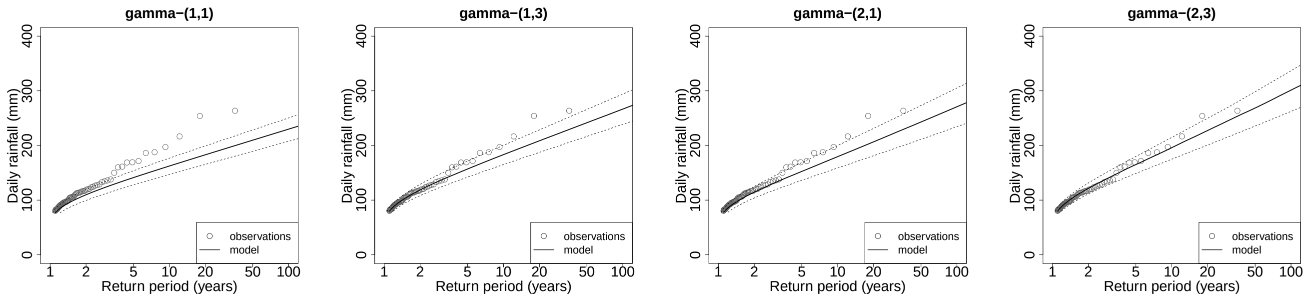

Figure 5: Case of Antraigues obtained when split sampling blocks of five consecutive days, when $G_{s,k}$ are Gamma distributions and the number of seasons and WT varies: $S \in \{1, 2\}$ and $K \in \{1, 3\}$. The values of (S, K) are indicated in the title. The dotted lines show the 95%-envelope of return level estimates over the 100 subsamples. The plain line shows the median estimates. The gray points show the full sample (35 years). Each estimation is based on half of these points.

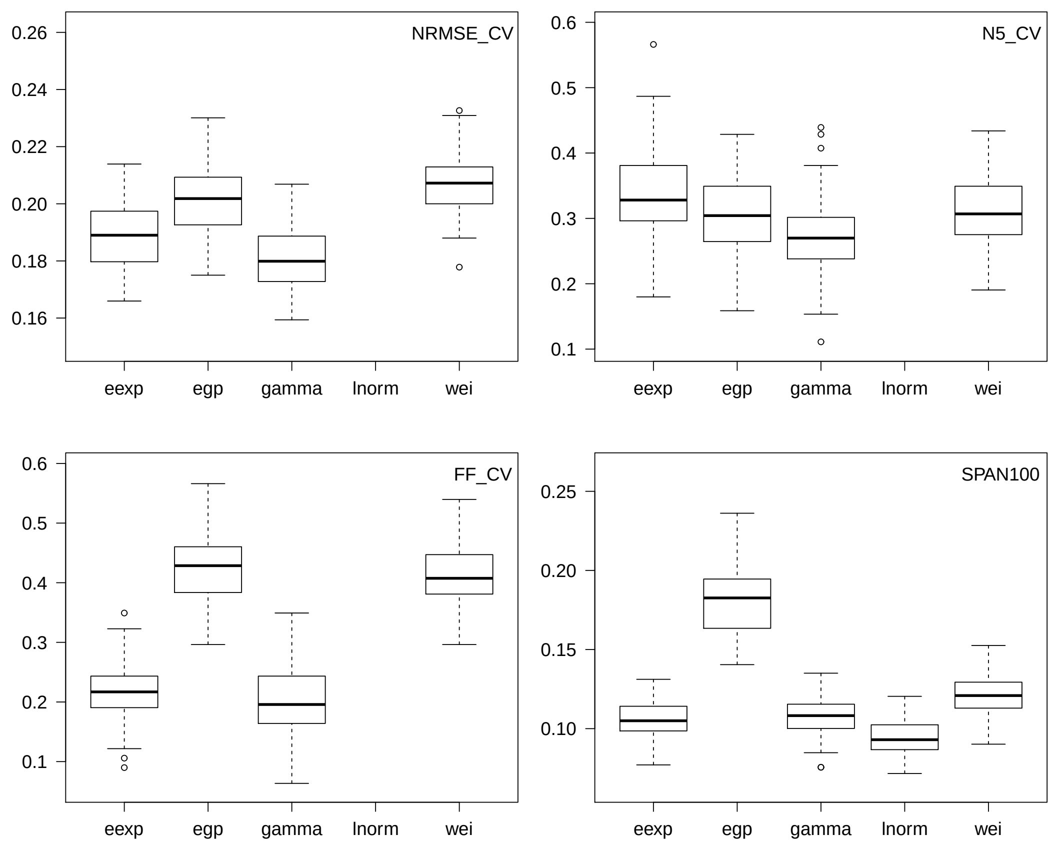

Figure 6: Scores of cross-validation obtained when split sampling block of five consecutive days, when $G_{s,k}$ is either the extended exponential (eexp), extended Generalized Pareto (egp), Gamma (gamma), lognormal (lnorm) or Weibull (wei) distribution, with $S = 2$ and $K = 3$. Each boxplot contains 100 points. The boxplots of reliability scores in the lognormal case are missing because they lie far above the upper range of depicted values.

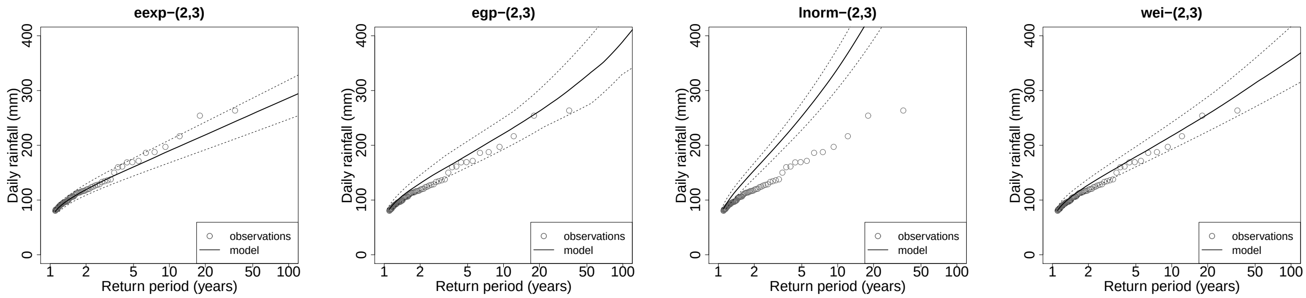

Figure 7: Case of Antraigues when when split sampling block of five consecutive days, when $G_{s,k}$ is either the extended exponential (eexp), extended Generalized Pareto (egp), lognormal (lnorm) or Weibull (wei) distribution, with $S = 2$ and $K = 3$. The dotted lines show the 95%-envelope of return level estimates over the 100 subsamples. The plain line shows the median estimates. The gray points show the full sample (35 years). Each estimation is based on half these points. Case of the Gamma distribution is shown in the right panel of Figure 5.

EDITORIAL COMMENTS

(1) p. 3, lines 16-17

Not clear how the "factor" is defined or calculated.

⇒ We apologize for the confusing formulation. By "a factor 1 to 2.6 is found for the annual totals of daily rainfall", we mean that the largest average of annual total of daily rainfall is 2.6 times larger than the lowest average of annual total (max= 2111 mm/year, min= 805 mm/year). To make it clearer, we propose to replace p. 3 lines 16-17 by: "... rainfall distribution. To illustrate these disparities, we show in Figure 2 the averages of annual totals and annual maximum daily rainfalls for each station. Computing the ratios between the largest and lowest values in Figure 2 gives a ratio of 2.6 for the annual totals and 3.2 for the annual maxima. For comparison the latter ratio is barely lower than the ratio found over the whole of France, which amounts 4. For both annual totals... "

(2) p. 25, Figure 10

Three of the graphs are for the same quantity, mean of non-zero rainfall for different weather patterns. But the color coding varies making comparisons difficult.

⇒ We show in Figure 10 the equivalent of Figure 10 of the article when using the same color scale for all the mean maps. It hinders visualizing the regional disparities in WP2 and 3. Therefore we prefer leaving Figure 10 of the article as it is.

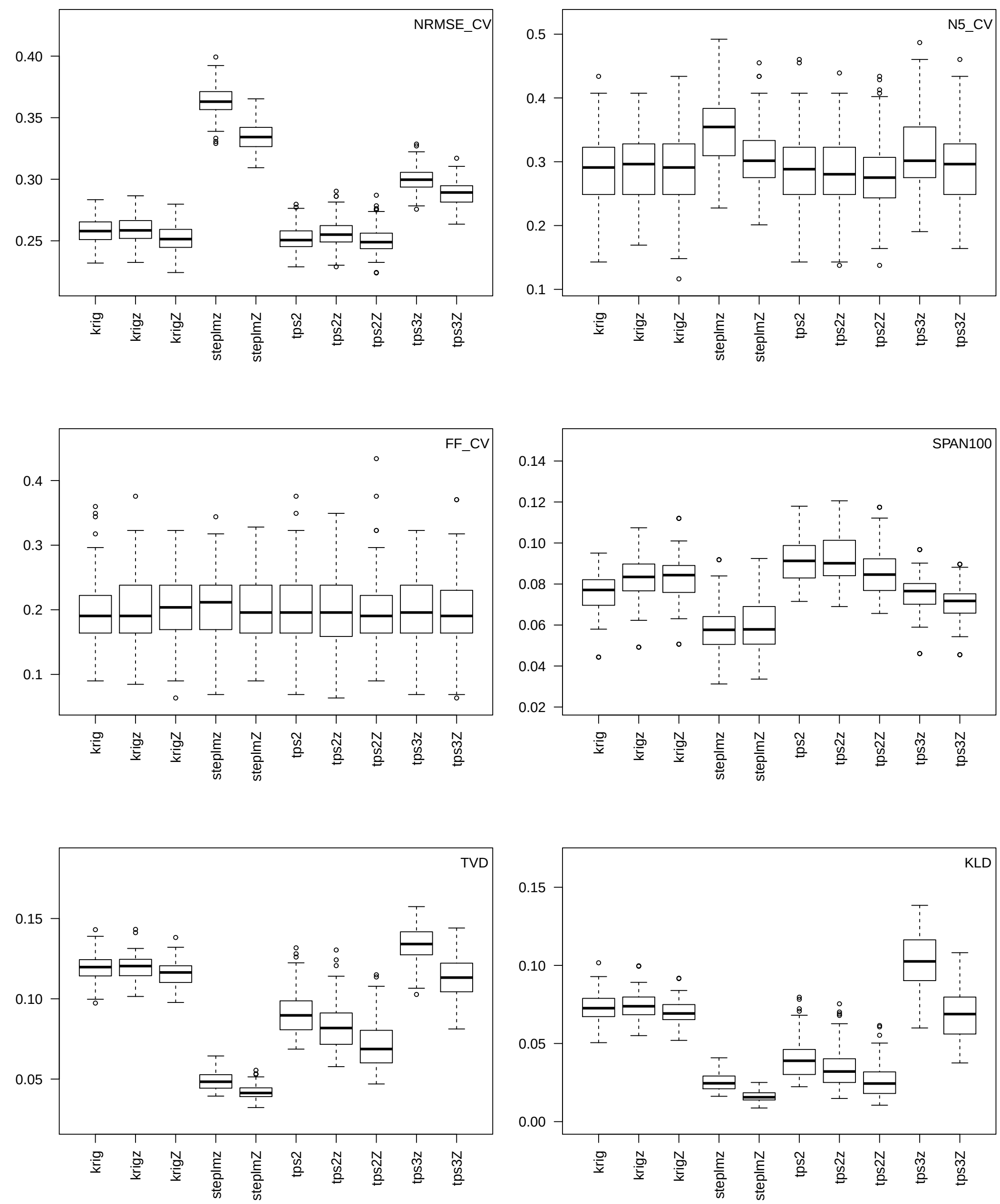

Figure 8: Scores of mapping obtained when split sampling block of five consecutive days, when $G_{s,k}$ are Gamma distributions with $S = 2$ and $K = 3$ whose parameters are interpolated with the mapping models of Table 3 of the article. The two first rows show leave-one-out cross-validation scores. Each boxplot contains 200 points. The third row compares interpolations at a given station whether the data of this station are used or not in the interpolation. Each boxplot contains 100 points.

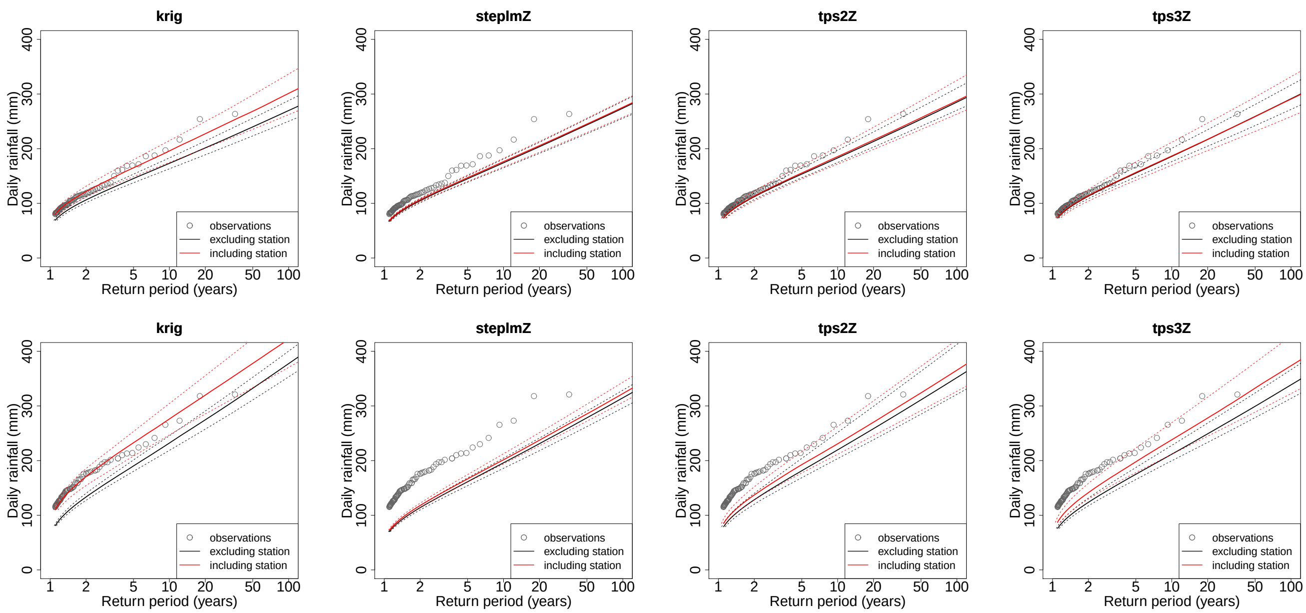

Figure 9: Case of Antraigues (top) and Mayres (bottom) when when split sampling block of five consecutive days, when $G_{s,k}$ are Gamma distributions with $S = 2$ and $K = 3$ whose parameters are interpolated with either: kriging without extrenal drift (krig), stepwise linear model (steplmZ), bivariate thin plate spline with drift (tps2Z), or trivariate thin plate spline (tps3Z). The dotted lines show the 95%-envelope of return level estimates over the 100 subsamples. The plain line shows the median estimates. In black, each interpolation is based on half the data of the other stations, excluding the considered station. In red, interpolation is based on half the data of all the stations, including the considered station. The gray points show the full sample (35 years for both stations).

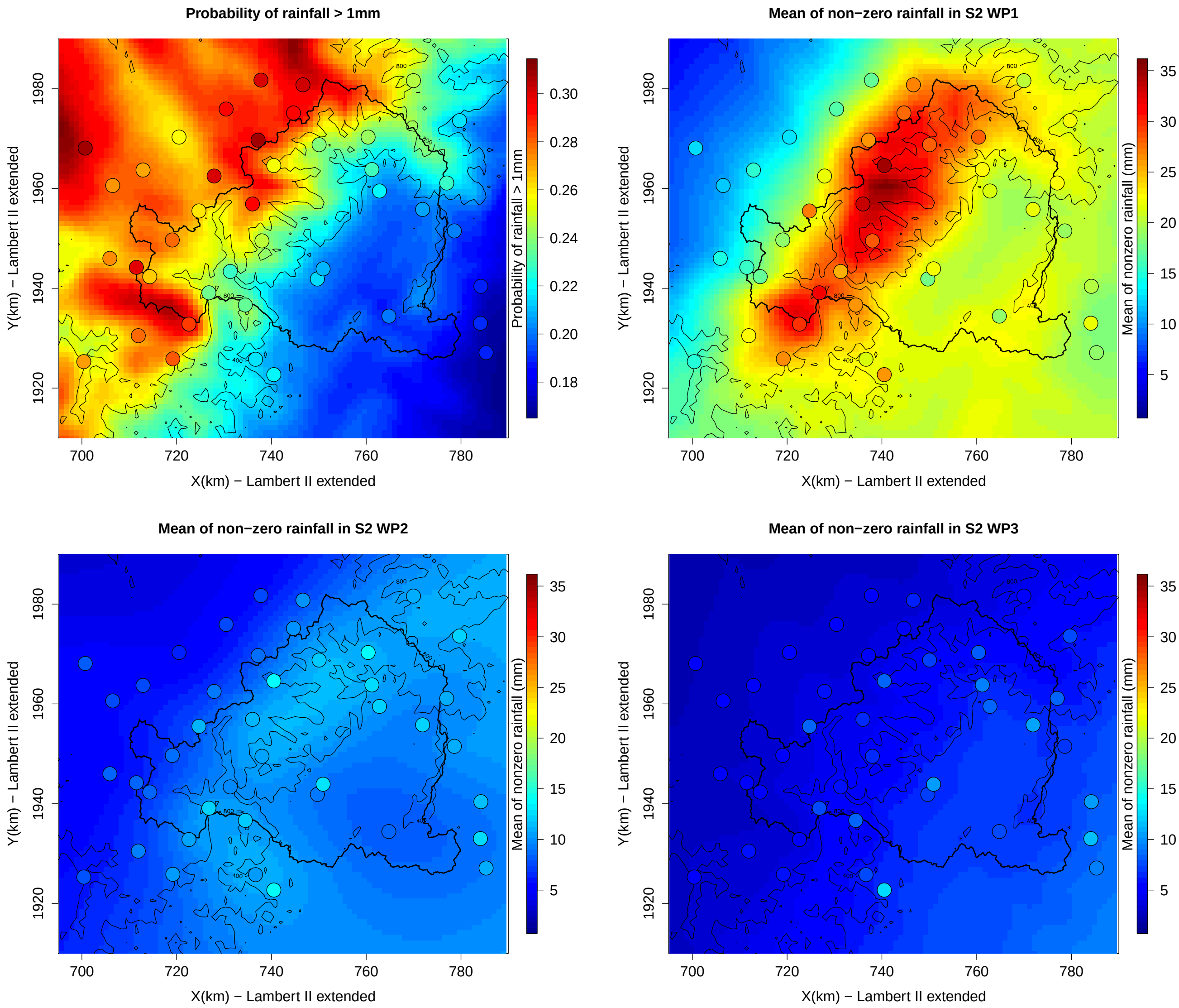

Figure 10: Map of the probability of daily rainfall to exceed 1mm and of the mean of nonzero rainfall in the three WPs of the season-at-risk. The points are colored with respect to the empirical estimates.

References

Blanchet, J., Touati, J., Lawrence, D., Garavaglia, F., and Paquet, E. (2015). Evaluation of a compound distribution based on weather patterns subsampling for extreme rainfall in Norway. *Natural Hazards and Earth System Sciences*, 15(12):2653–2667.

Brigode, P., Bernardara, P., Paquet, E., Gailhard, J., Garavaglia, F., Merz, R., Mićović, Z., Lawrence, D., and Ribstein, P. (2014). Sensitivity analysis of SCHADEX extreme flood estimations to observed hydrometeorological variability. *Water Resources Research*, 50(1):353–370.

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

Ferro, C. A. T. and Segers, J. (2003). Inference for Clusters of Extreme Values. *Journal of the Royal Statistical Society. Series B (Statistical Methodology)*, 65(2):pp. 545–556.

Frigessi, A., Haug, O., and Rue, H. (2002). A dynamic mixture model for unsupervised tail estimation without threshold selection. *Extremes*, 5(3):219–235.

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

Garavaglia, F., Lang, M., Paquet, E., Gailhard, J., Garçon, R., and Renard, B. (2011). Reliability and robustness of rainfall compound distribution model based on weather pattern sub-sampling. *Hydrology and Earth System Sciences*, 15(2):519–532.

Gottardi, F., Obled, C., Gailhard, J., and Paquet, E. (2012). Statistical reanalysis of precipitation fields based on ground network data and weather patterns: Application over French mountains. *Journal of Hydrology*, 432–433(0):154 – 167.

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

Scarrot, C. and MacDonald, A. (2012). A review of extreme value threshold estimation and uncertainty quantification. *REVSTAT - Statistical Journal*, 10(1):33–60.