We warmly thank the Reviewers for their valuable comments. We provide below a detailed point-to-point reply to these comments. The changes included in the new version of the article are indicated in red.

1 Answer to Reviewer #1

GENERAL COMMENTS

The revised version of the manuscript is somewhat improved, as the authors have addressed my comments at least to some extent. My main remaining reservation concerns how the results on the fitting of marginal distributions are interpreted in terms of heavy tails.

For this reason, I recommend that the manuscript be accepted for publication subject to minor revision.

SPECIFIC COMMENTS

(1) Interpretation of results concerning heavy tails

The interpretation of the results concerning the shape of the upper tail of the distribution of rainfall amounts still suffers from a lack of reliance on what is known from extreme value theory. In particular, the return level plots for the fitted mixtures of gamma distributions (Fig. 7) all appear approximately linear for high values (i.e., indicative of light-tailed, not heavy-tailed, distributions). Apparently, using a mixture of gamma distributions reduces the bias attributable to fitting the entire span of the values of rainfall rather than only fitting the upper tail (i.e., the mixture of gamma distributions does not necessarily have a heavy, or even a heavier, tail in the sense of extreme value theory).

Along the same lines, the corresponding return level plots (Fig. 9) for sub-exponential distributions with heavy tails (in either an asymptotic or sub-asymptotic extreme value theory sense) do not necessarily indicate that the fitted models have too heavy a tail. Rather, their lack of fit in the upper tail is at least partially attributable to being based on fitting the entire range of rainfall values.

For this reason, I fear that the statement in the Conclusion and discussion section "distributions showing sub-exponential tails (EGP for example) give usually unrealistic return levels" (p. 28, line 18) will be misunderstood. At best, it is only relevant to the situation in which the distribution (or mixture of distributions) is being fitted to the entire range of rainfall values. As such, it does not necessarily imply anything about how heavy the upper tail of the distribution of rainfall amounts actually is.

⇒ We apologize for the confusion. We agree that, although the mixture of Gamma distribution allows to get larger return levels than the Gamma distribution alone, it is still unable to produce heavy tails in the sense of extreme value theory. We have made this clearer in the first paragraph p. 20 section 4.1, that has been partly rewritten, stating in particular : "However we note that the return level plots of Figure 7 all appear approximately linear for high values, meaning that none of the Gamma mixtures is able to produce heavy tails in the sense of extreme value theory. It is possible that return levels at extrapolation far beyond the observed return periods are underestimated." We also recall this in the last paragraph of Section 5 p. 28, that has been partly rewritten, stating in particular: "Possible direction of improvement for the study region regards the choice of the marginal distribution. Although the Gamma mixture was selected according to the cross-validation scores, we noted a possible underestimation of return levels at far extrapolation since the model is unable to produce heavy tails in the sense of extreme value theory."

We also agree that the lack of reliability of the extended Generalized Pareto distribution might be partly attributable to fitting the entire range of values. In particular, we state in the article that the FF and N_T scores tend to be too frequently too small with the extended Generalized Pareto distribution (see Section 4.1), implying a systematic overestimation of the probability of occurrence of large values. This leads to very large return level estimates associated to large return periods. Figure 1 shows that, although both models are theoretically equivalent for extreme values (Naveau et al., 2016), the extended Generalized Pareto gives much larger return level estimates than the Generalized Pareto distribution, due to an overestimation of the shape parameter. The median values of the 500-year return levels exceeds 550 mm/day. This is larger than the largest return level with the Generalized Pareto distribution. We have made this clearer by removing the corresponding sentence and adding in the second paragraph of p. 20 section 4.1: "Note that the lack of reliability of the extended Generalized Pareto in the upper tail is at least partially attributable to being based on fitting the entire range of rainfall values, which leads to a systematic overestimation of the shape parameter ξ in Table 1 compared to when fitting a Generalized Pareto distribution on the upper tail of the data (not shown)." Note also that we considered here the most simple version of the extended Generalized Pareto distribution, while other versions may possibly show better performance (see the second paragraph of p. 28 Section 5: "Other possibility includes considering less parcimonious versions of the extended Generalized Pareto distribution (Naveau et al., 2016) to improve reliability in the upper tail.").

EDITORIAL COMMENTS

(1) Table 1

The table indicates that CDFs are listed. Yet for some distributions (i.e., gamma and lognormal) the expressions given are actually for probability density functions, not CDFs.

⇒ Actually Table 1 indicates: "CDF $G(r)$ or density $g(r)$".

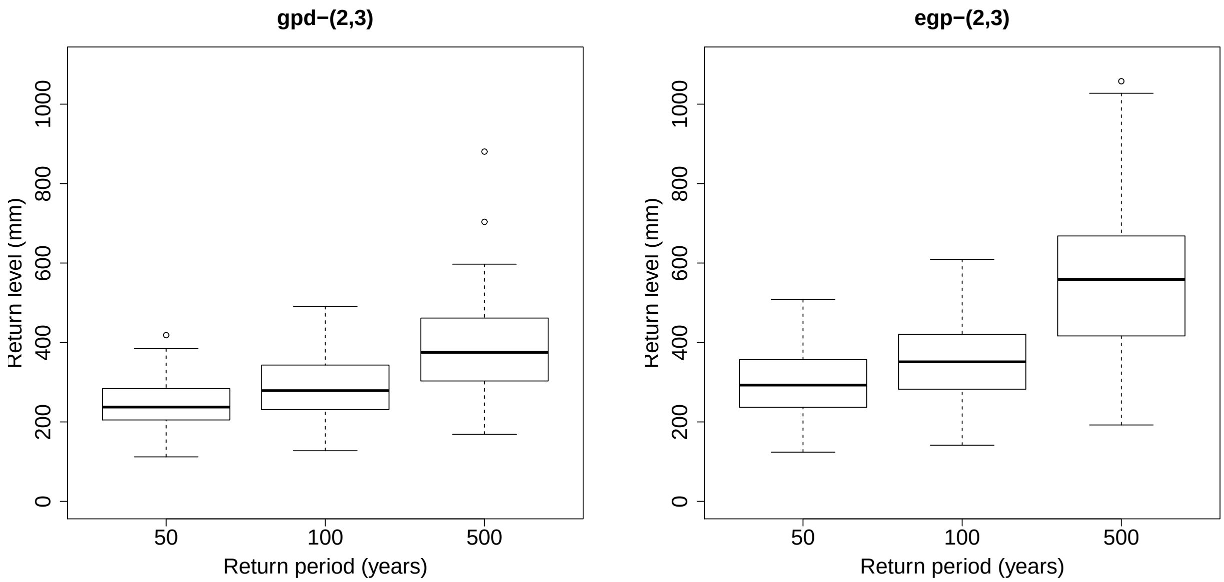

Figure 1: Boxplot of the return levels estimates of the 42 stations with (left) the mixture of Generalized Pareto distribution fitted to the 90%-quantile exceedances in each of the 6 subclasses, (right) the mixture of extended Generalized Pareto distributions. Estimations are made on the full data.

2 Answer to Reviewer #2

The manuscript proposes an objective framework for mapping rainfall hazard in an area. It aims at both evaluating the best statistical distribution at station location and assessing the best mapping method. The authors propose the adoption of a unique distribution for the full distribution paying a particular attention to the tails. A small catchment in France is used for assessing the potentialities of the proposed techniques. The manuscript significantly improved compared to the first version, but I think some of the key-points underlined from both Reviewer#1 and Reviewer#2 are still open, and require to be more carefully assessed for making the manuscript ready for publication.

MAJOR COMMENT

The major issue is related to the aim of the manuscript. In the answers to the reviewers (R#1 answer 1, R#2 answer 1), and in the manuscript itself (e.g., P1L1), the authors declare that the aim of their work is not to assess the best probability distribution or mapping method in the Ardèche catchment they are analysing, but to propose an effective operational framework for choosing the best approach for mapping rainfall hazard.

Despite this, the results and conclusions sections are not focused on assessing the efficiency of the method in discriminating and choosing the best distribution-mapping method pair for a certain region, but on the description of the results for the case study. As the case study is related to a unique region, and that no-other selection approach or regional study is proposed as comparison and feedback, I can not understand how the *author can assess the efficiency of the proposed framework.*

⇒ We thank the Reviewer for this comment. However we think there is a misunderstanding here. Sections 3.1.2 and 3.2.2 are entirely devoted to the introduction of scores aiming to objectively compare marginal models and mapping methods (based on their reliability and stability). Sections 4.1 and 4.2 of the results (7 pages in total) use these scores to choose the best distribution-mapping method pair for the region. Only Section 4.3 (one page text + two figures) illustrates the selected model for the region. Of course, we could remove Section 4.3 since the main goal of the article is the cross-validation framework but we find it interesting to briefly illustrate some rainfall features from the selected model, although this is not the main scope of the article.

I strongly agree with R#2, when he says (answer1) that the authors should do a better work in underlying and demonstrating that the proposed framework is a really significant step forward compared to the state-of-the-art in hazard-mapping. The authors are stating their aim is not to propose a new way of mapping rainfall hazard, but at least, considering that the proposed framework is not particularly innovative from the methodological point of view (as it consists in chaining two standard cross-validation procedures using some regional statistics to verify the combination that provides the best score), I think it's quite important to analyse the common approach adopted for this operation in the literature, and assess the improvement provided by the author's methodology.

⇒ We agree that the idea of fitting a marginal distribution and interpolating the parameters is not new and we have never claimed so. However we are not aware of any study comparing objectively the full procedure going from point measurements to mapped distributions. This is the gap this article intends to fill. We are sorry to admit that we are not aware of the "common approach adopted for this operation in the literature". To the best of our knowledge and as stated in the introduction, several studies have compared marginal distributions, other have compared mapping models, but there is nothing regarding the full procedure. Furthermore, in the great majority of the studies, the scores of comparison are NRMSE, which is also one of our scores but we complement this with several other scores designed for extreme values or/and assessing robustness and stability.

Even the effects of one of the stronger assumption adopted from the author in their methodology (the use of an unique function for the full distribution) are not effectively assessed and just based on some consideration derived from the literature (R#1 answer2).

⇒ We fully agree that using a unique distribution for the full spectrum of values is questionable, as also stated at the very top of p. 6. However let us first point out that in this article mixtures of (single) distributions (over seasons and WPs) are considered. As shown in the results section, this allows a much greater flexibility and reliability than using a unique distribution without mixing, in particular for the tail.

Another possibility would of course be to use hybrid models based on combining distributions for low and heavy amounts. Very honestly, we chose not to consider hybrid models for the sake of concision since many hybrids models have been proposed in the literature (Vrac and Naveau, 2007; Furrer and Katz, 2008; Li et al., 2012) and there seems to be no consensus on which one is to be preferred. However let us recall that the considered distributions/mapping models only aim at illustrating the proposed framework of model selection. We do not think that multiplying the considered distributions/mapping methods would be much helpful in this case. However we fully agree that if one would like to select the best model for his data, hybrid models would definitively for worth of consideration. We have made this clearer in the conclusion section p. 28: "It could be worth considering hydrid models based on combining distributions for low and heavy amounts (Vrac and Naveau, 2007; Furrer and Katz, 2008; Li et al., 2012), although robustness might be an issue." We have also modified the title and partly rewritten the Abstract and the first paragraph of the conclusion in order to better stress that the scope of the paper is the cross-validation framework.

Concluding, I think the authors should more strongly stress the improvements that their framework and the hypothesis they set, can provide. This could be done by comparing it with the classical "not coupled" methodologies, commonly adopted for selecting separately the best distribution and the best mapping method or trying to verify the effect of different configurations of the framework (e.g., adopting hybrid distributions) on the results on their case study. If it is not feasible, they should at least test the technique on other basins, to provide evidence of the ability of the technique to effectively distinguish the best distribution-mapping technique pairs, according to the different characteristics of the basin.

⇒ We thank the reviewer for this comment but we are not sure to understand what is meant by "the classical not coupled methodologies". Here the methodology is also not coupled since we first select the marginal distribution, and then we select the mapping method based on the selected marginal distribution (see Figure 5 of the article). Regarding the idea of testing the framework on other basins, of course it is of interest. Actually, we have also applied the cross-validation framework on a much larger basin, the Durance basin (14,000 km^2), in South of France. However we did not present the results here because the Ardèche basin is only chosen for illustration purpose, as a way of showing how the framework works and how it allows to make selection of the best model. To make this clearer, we have slightly modified the abstract, conclusion and the title.

MINOR ISSUES

I think the manuscript still require a final proofread by a native speaker as a number of language mistakes still arises. E.g.:

- P1L6 – I think "inhomogeneities" is more appropriate than "disparities" when refers to spatial distributions.

⇒ Done.

- *P1L18 - "fields" → "field"*

⇒ Done.

- *P2L4 – I don't think "learning" is the correct word.*

⇒ Replaced by "estimating".

- *P2L11 – "methods are able" → "method is able"*

⇒ Done.

- *P3L3 – "lower elevated Rhnoe Valley" → Incorrect, please rephrase*

Rightarrow "lower elevated Rhône valley" replaced by "Rhône valley".

- *P5L6 – "they are not be considered" → "they are not considered"*

⇒ Done.

- *P6L19 – "distribution is an unsupervised way" → "distribution in an unsupervised way"*

⇒ Done.

- *P9L11 – I don't think "#" can be used as variable.*

⇒ We have replaced the notation by card(B_c).

- *P14L9 and across the manuscript – "450mm... 1mm" separate the measurement units from the number "450 mm... 1 mm"*

⇒ Done.

References

Furrer, E. M. and Katz, R. W. (2008). Improving the simulation of extreme precipitation events by stochastic weather generators. *Water Resources Research*, W12439.

Li, C., Singh, V. P., and Mishra, A. K. (2012). Simulation of the entire range of daily precipitation using a hybrid probability distribution. *Water Resources Research*, 48(3). W03521.

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.*

Vrac, M. and Naveau, P. (2007). Stochastic downscaling of precipitation: From dry events to heavy rainfalls. *Water Resources Research*, 43(7). W07402.