

Dear reviewer,

Thank you for taking the time to carefully review our manuscript, and thank you for your various comments. The latter will help us bring greater clarity and scientific rigor to our study and improve its overall quality.

First of all, when running the sensitivity analysis, we found an error in the dataset of amax we selected for the study. We repeated the analysis running different set of parameters for the analysis, and the one we selected turned out to be looser than expected. With this more selective dataset, we found 71 stations rather than 80. The conclusions that we found were similar regarding the regional majority of non stationarity and the temporal patterns. We only found 5 clusters. There is no more the precedent clusters 'Mixture' and 'Stationary' which included a few stations with missing data (removed in the updated results), but clearer patterns found in the updated Clusters.

We thank you for the suggestion to test multiple datasets that helped us find an error and refine our selection.

We have addressed each comment below. Reviewers comments are shown in bold, line by line, followed by our response.

Lines 25-36

We will be more concise in our contextual background and get straightforward to the relevance of the study and to the need to work on hydrological extremes.

Lines 37-48

We will be sure to clearly make the distinction between hazard, exposure and vulnerability in the paper. There is an increase in the hydrological risk (di Baldassarre et al. 2010) in West Africa. Flood hazard can be a potential factor involved in the increase of the hydrological risk. But the lack of dense and long-term data and the difficulty to detect hydrological signals due to hydrological variability both retrain the possibility to assess this potential influence. However, some studies have shown, locally, an increase in flood hazards. We clarify that the aim of our paper is to investigate if there in a regional increase in flood hazards.

Lines 87-107

The paper suggests references to important papers in litterature about distinction of typical hydrological behaviors in West Africa.

We are conscious of the lack of clarity or ambiguity of the concepts, but there is neither predefined assignment nor a spatial criterion actually existing for each watershed. Studies to which we refer suggests concepts which are still actually vague. We will be more explicit about that.

Lines 116-125

Thanks for the remark for the need of clarifying.

We will provide the different parameters of the maxima extraction in Supplementary material.

The climatological hydrograph was constructed this way :

For each Julian day J, we average discharge values over the 75 years of data, provided we have at least 10 years of data. We obtain a nearly complete hydrograph, with variability smoothed out

thanks to the large number of observations. If there are missing days in the hydrograph, we interpolate them.

From there, we determine the highest peak (absolute maximum) and identify its base (left and right points). If there are two peaks or not a clear peak, this selection method allow to get a larger window, encompassing the period were the hydrograph is high.

Lines 124-129

We thank you for the suggestion and we will provide the retained AMAX series under different criterion of missing data.

Lines 126-133

The distribution of valid years and of the annual maxima captured at each station will be summarized in a heatmap. We acknowledge that we only have series from stations with great monitoring continuity, but it is in line with the scope of robust trend detection, requiring dense and long-term data. Lack of data could lead to misrepresentation of temporal patterns. The choices made for defining the criteria were a compromise, prioritizing the exigency of detection of trends. We could propose an analysis with series with only stricter criteria.

Lines 135-138

Various elements will be made available, such as the AMAX series calculated for the selected network.

AMAX series from the ADHI database are already available here <https://dataverse.ird.fr/file.xhtml?fileId=4921&version=6.1>. The parameters of the fitted model will also be made available in a github.

Lines 165-178

We have adjusted one breakpoint model year by year, and we have selected the best model among all these models.

We simply exclude the first and last years of the observation period because this would correspond to an almost entirely linear trend over most of the time window. As a boundary effet, the remaining portion of the trend could not be considered robust as it would be calculated based on only a few years.

Finally, the model does not impose a specific timing, as several breaking years are proposed; it simply selects the most appropriate one from among the various proposed break years. If this pattern is not valid, this model will not be selected anyway.

Lines 175-178:

We plan to add a paragraph in the Discussion about the treatment of the scale parameter. We considered the use of $\sigma(t)/\mu(t) = \text{constant}$, based on previous studies in West Africa (Wilcox et al. 2018).

Lines 180-218

We understand the problem of the potential influence of the order of comparisons. But, the process would require a significant number of stationary simulations. We consider the methodology very expensive compared to the aim of finding a order of magnitude of stationay stations.

However, we have tested different choices of selection, such as direct AIC and BIC criteria, and we will provide them in the future manuscript.

Lines 182-192

We did not understand the question. If we understand correctly, we did not apply the likelihood ratio test with the BreakPoint model, as it is not embedded in any other one. We thank you for the commentary but we would appreciate a clarification of it.

Lines 219-223 and Lines 225-228

We will complete the analysis with the estimation of the 10-year return levels. However, we think it will barely change the results as we work with normalized return levels. We do not think it is appropriate to compute return levels larger than 10 years as we only have 35 years of observations and we are interested in trends.

Lines 229-241

We will put the described parameters for the clustering in the Appendix in order to improve the reproducibility and interpretation.

We decided to plot the result for the window 1950-2020 (rounded). We could have chosen 2018, but the difference between the two would be insignificant as we work on normalized return levels.

Lines 245-268

We understand the comment, but we think that the bootstrap is actually out of the scope. The estimation of the different models for all stations is about 3 hours. Doing 1000 bootstraps the computation time would approximate the $3\text{hours} \times 1000 = 125$ days. As a reminder, the scope is to get a regional overview of flood trends, not a specific basin-scale analysis. However, we will remind the uncertainty of model selection in the paper.

Lines 270-276

The $k=6$ were retained for interpretability (this k will change in new results). It was a guided choice after several tests. We will show the different classification with different k in supplementary materials.

Lines 277-296

We thank you for the comment and we will clarify interpretation of the Upward cluster in the description of the clusters, by distinguishing the different regional evolution with a more local evolution. We will also discuss local hydrological behavior more deeply adding references.

Lines 284-291

We will reformulate to assess the results are analyzed using visual interpretation.

Lines 297-304

- We were not looking for strong evidence of hydrological controls. A further analysis could have been lead if we have had strong indices. As there are stationary features, we just want to get an overview of the description.
- These are the same discharge records used for the computation of hydrological indices, but the information retained in clusters is about the temporal evolution pattern of maxima,

which is different information from the hydrological indices. However, we will precise and better insist on this and the fact that the indices do not intend to be explanatory or causal.

Lines 305-314

We will better phrase this paragraph and we will provide a table in Supplementary materials with discriminant features retained for each cluster.

Lines 318-335

We will rephrase and highlight it the way you suggested ; non-stationary models are frequently selected among the retained stations under the proposed modelling framework, which supports the findings of previous studies – on a much larger regional scale – regarding the regional non-stationarity of floods in West Africa.

Lines 351-356

We will rephrase this part explaining that the comparison is not direct, as we do not work on the same data neither the same methods.

Lines 363-380

We will better phrase this part of the Discussion, explaining that we formulate hypotheses rather than we suggest spatial associations. As we explained earlier, we can not provide an explicit distinction of hydrological behaviors as there is not specific assignment nor spatial criterion of this distinction. This analysis could have provided suggestions for attribution pathways, but it is not the case.

Lines 397-409

We will put the comparison with AIC and BIC in the Supplementary material as we suggested earlier, with the different changes in model selection.

Lines 401-409

We thank you for the suggestion and we will add the reference in the paper.

Lines 410-415

We will put the necessary material in Supplementary material in order to show the structure of the clusters.

Lines 418-445

We agree and we will rephrase these sentences to avoid any overly broad interpretation.

We thank you again for your comments, and we hope that the upcoming corrections and the revised version will meet your expectations.