**The manuscript titled “Changes in Mediterranean flood processes and seasonality” by Tramblay et al. investigates how flood types and seasonality evolve in Southern France during the past 50 years and attempts to link the change in flood seasonality to changes in mechanisms. Overall, the logic is clear and the results are well presented. However, deeper scrutiny reveals some issues that, in my opinion, might undermine the quality of the manuscript in its current form. My main concerns are the robustness of the results and the contribution of the study, which are relatively weak in the present manuscript.**

We would like to thank you for the review of our paper and the constructive comments to improve the presentation of the results.

**Firstly, in the introduction part, the authors argue that “Most of these studies rely on flood classification schemes, with various complexity depending on the type of data available, allowing a data-based separation of floods into their distinct generation mechanisms”, which I agreed with. While I assumed that the study would propose an improved approach to partially overcome the current limitations of the flood classification scheme (e.g., the relatively subjective threshold selection, etc.), it seems that the study only used a simple decision tree with hard thresholds, without any justification or discussion of the threshold selection (Tarasova et al. 2020; Zhang et al, 2022). Currently, all the results are based on the seemingly arbitrary threshold and structure of the decision tree. For example, the authors use “50% saturation”, “95th percentile of rainfall”, etc. as the threshold to distinguish the “excess rainfall”, “long rainfall”, and “short rainfall”, while if some events are distributed around these critical points, the conclusion about their changes might be quite sensitive to these values. I wonder if there are some sensitivity tests to ensure the robustness of the results.**

The statement made by the reviewer: “Currently, all the results are based on the seemingly arbitrary threshold” is not correct. The results presented in section 4.1, about changes in flood events characteristics, in section 4.2 about changes in flood dates and in section 4.3 about the association between flood occurrence and weather patterns <u>do not depend upon the flood event classification</u>. Only results from the sections 4.3 and 4.4 are using the flood classification.

We agree with the reviewer that we did not provide enough information about the sensitivity of the results to the threshold values used in the classification. We want to stress that we do not aim in the present study to introduce an improved approach to classify floods events. Instead, we provide a regional focus in the Mediterranean, where floods have very strong impacts, using a well-used classification (to allow inter-regional comparisons) applied globally, in Stein et al. (2020), over the Continental USA in Stein et al. (2021) and over Africa by Tramblay et al. (2022). We indeed used very basic thresholds, yet adapted to the processes analyzed: the 95$^{th}$ percentiles for extreme precipitation and a threshold of 50% for the soil wetness index to define wet/dry conditions (these thresholds, relevant for these processes, are widely used see a recent example in Nanditha, J. S., & Mishra et al. 2022).

Following the reviewer advice, we implemented a Monte Carlo experiment, based on Latin Hypercube sampling to analyze the sensitivity of the regional changes in flood types to the thresholds used in the classification (using 5000 experiments with these threshold ranges: Extreme rainfall [0.7 1], Soil Wetness Index [0.4 1]). The question we want to address here is: <u>are the changes we detected in the regional flood types dependent on the thresholds used in the classification?</u> See the boxplot below, <u>the response is no</u>, we have a decrease in excess rain floods and an increase in short rain floods that are not dependent on classification thresholds.

On the opposite, for long rain floods, we can see that the confidence interval includes zero, so the changes are not robust. There are indeed interplays between the “excess-rain” and “long-rain” categories, with the soil moisture threshold increasing, less events are classified into “excess-rain” and more are classified into “long-rain”. The results of the sensitivity analysis have been included in the revised manuscript.

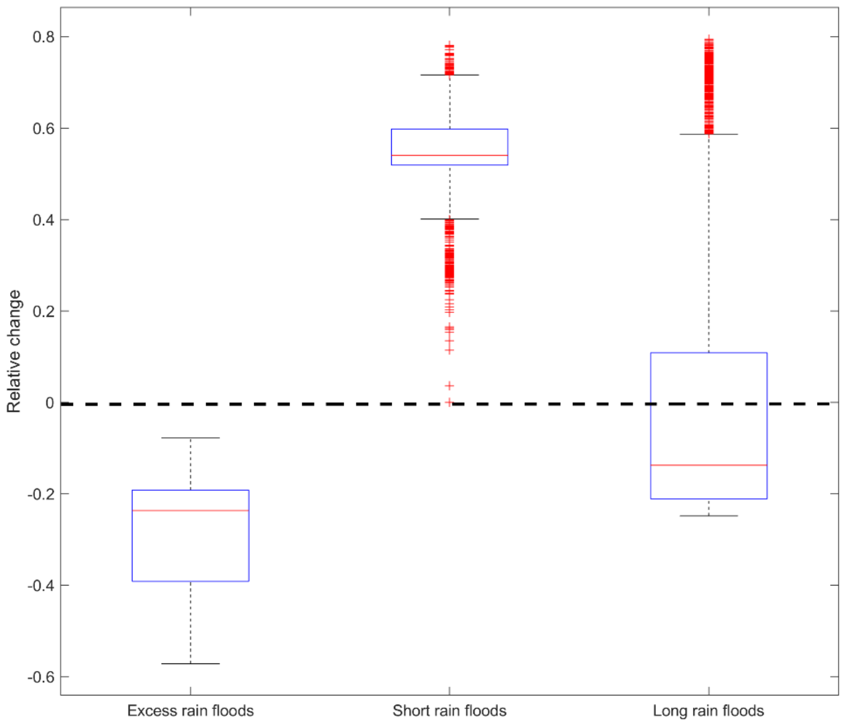



Figure 1: Relative changes in the regional frequency of excess rain floods, short rain floods and long rain floods, given a range of classification thresholds analyzed via a Monte Carlo experiment. For each box, the central mark indicates the median, and the bottom and top edges of the box indicate the 25th and 75th percentiles, respectively. The whiskers extend to the 99% of the data and the 1% outliers are plotted with the '+' marker symbol.


Nanditha, J. S., & Mishra, V. (2022). Multiday precipitation is a prominent driver of floods in Indian river basins. Water Resources Research, 58, e2022WR032723. https://doi.org/10.1029/2022WR032723


**In addition, the manuscript argued that “Yet, beside the trend detection, no study has provided an in-depth analysis of the long-term evolution of flood processes in these regions”. This seems to be a very strong statement. However, a recent study by Jiang et al. (2022) has examined the change in flood mechanism in 1000 catchments in Europe, which also include catchments in southern France. The study showed similar trends in flood generation processes as in the present manuscript. I think it would make more sense to compare the trend results with other literature.**

We agree and we modified the statement to be more specific: “Yet, beside trend detection or changes in flood types, no study has provided an in-depth analysis of the long-term evolution of flood processes in Mediterranean basins, in relation to their drivers such as precipitation, soil moisture and the evolution of synoptic weather patterns associated with floods”.

We acknowledge that it is important to compare results in different studies to see if different approaches can provide similar conclusions. Besides large-scale studies, it is also interesting to have a deeper look at the regional level with good quality datasets. For instance, the European dataset EOBS over France includes about 120 meteorological gauges, and the SIM product, used in the present study, over 4000. It is very important to document if the same types of relationships are observed across different scales and different databases, notably to assess the relevance of these large-scale datasets in data-sparse regions.

The reviewer seems to consider that "the contribution of the study is weak", notably in relation to the recent work of Jiang et al. 2022. The study of Jiang et al, 2022 is mostly centered over Central and Northern Europe, using EOBS precipitation and temperature and river discharge. They include about 38 basins in South France (from their map), 13 in the Iberian Peninsula and 1 in Italy, south of the Po basin, so for a total of 1000 basins, Mediterranean basins represent 5.2% of the basins studied. It is interesting to note that we obtained similar results about the reduction of Excess rainfall floods (as in Tarasova et al., 2023), but Jiang et al. (2022) did not consider changes in flood seasonality, flood magnitude depending on event types, or flood characteristics such as runoff coefficients, as we did in the present study.

We added in the introduction: "Recent large-scale studies (Jiang et al., 2022, Tarasova et al., 2023) suggested a reduction of flood driven by soil saturation, including basins in the Mediterranean area."

**Also, my concerns about the contribution arise because the conclusion is not very strong currently based on the significance test results. Although the authors argued that it is due to the short records of samples and interannual variability, it would impair the reliability of the conclusion somehow, particularly in the case of lack of sensitivity test for the method.**

About significance testing, as shown on figure 11 the regional changes in the frequency of flood event types are significant for the reduction of excess rain floods and the increase in short rain floods. It should be reminded here that we are dealing with extreme events in the Mediterranean hydro-climatic context, with a strong year-to-year variability. Therefore, at the basin scale, dealing with a small number of extreme events, it would be quite difficult to see 'statistically significant' changes given the small sample sizes. More, a multiplication of local tests is not the best approach (Wilks, 2016). This is why we adopted a regional analysis to check the significance of the regional changes. As noted above, we also included in the revised manuscript a sensitivity analysis to the threshold values.

In addition, to put more strength on the results, since the main question is whether the detected changes are regionally significant, we also added two additional regional significance tests:

1. To assess the regional changes in flood events characteristics (the total and extreme rainfall, runoff coefficients, contribution of baseflow, flood duration and soil moisture), we performed a regional pooling of the events and applied the Mann-Kendall test to detect trends in the regional series of event characteristics. As shown in the table below, all the detected changes are regionally significant except the decrease in base flow contribution to peak discharge during floods. We added this result in the revised manuscript.

| Flood characteristics | pvalue of the regional MK test | Regional changes (%) |
|---|---|---|
| Flood event duration | **0.0046178** | -0.40% |
| Base flow contribution to peak | **0.5687962** | -8.62% |
| Runoff coefficient | **0.0000002** | -14.62% |
| Total event rainfall | **0.0011851** | 9.01% |
| Maximum event rainfall | **0.0000000** | 13.47% |
| Antecedent soil moisture | **0.0000008** | -9.80% |

2. To assess the regional changes in flood dates, we first separated in two regional samples the stations where floods tend to occur earlier (sample 1) or later (sample 2). Then we used the Watson-William test, previously used to assess changes in flood dates in each station, to compare these two regional samples. The test results indicate that for the 19 stations where floods tend to occur later, the change in flood dates are not significant at the 5% level (p value = 0.0821), on the opposite, for the 79 stations where floods are occurring earlier, the change is significant (p value = 5.34.10^-8).

We added these results in the revised manuscript.

Wilks, D. S. (2016). "The Stippling Shows Statistically Significant Grid Points": How Research Results are Routinely Overstated and Overinterpreted, and What to Do about It. In Bulletin of the American Meteorological Society (Vol. 97, Issue 12, pp. 2263–2273). American Meteorological Society. https://doi.org/10.1175/bams-d-15-00267.1

**References:**

**Tarasova, L., et al. A process-based framework to characterize and classify runoff events: The event typology of Germany. Water Resources Research, 56, e2019WR026951 (2020). https://doi.org/10.1029/2019WR026951**

**Zhang, S., et al. Reconciling disagreement on global river flood changes in a warming climate. Nat. Clim. Chang. 12, 1160–1167 (2022). https://doi.org/10.1038/s41558-022-01539-7**

**Jiang, S., et al. River flooding mechanisms and their changes in Europe revealed by explainable machine learning, Hydrol. Earth Syst. Sci., 26, 6339–6359 (2022), https://doi.org/10.5194/hess-26-6339-2022.**

**Other comments follow:**

**Abstract: what do the "flood event characteristics" mean? Please specify.**

We added: "Further, several flood events characteristics have been computed: flood event durations, base flow contribution to the peak, runoff coefficient, total and maximum event rainfall and antecedent soil moisture".

**L131: What is the spatial resolution of the SIM reanalysis data used? Please clarify if the size of the study catchments is comparable to the spatial resolution of the hydrometeorological data sets.**

The spatial resolution is 8x8km (we added this information, that was missing, in the revised manuscript). SIM is the best available product based on observations covering France, since 1958. 83% of basins considered in the study have a size larger than 64km², the mean catchment size considered in the present work is 480 km².

**L133: Please clarify how the nival regime is identified from the river discharge hydrographs. It is also not clear why they should be removed, since even snow-covered catchments can also be affected by rainfall, e.g. rain-on-snow events.**

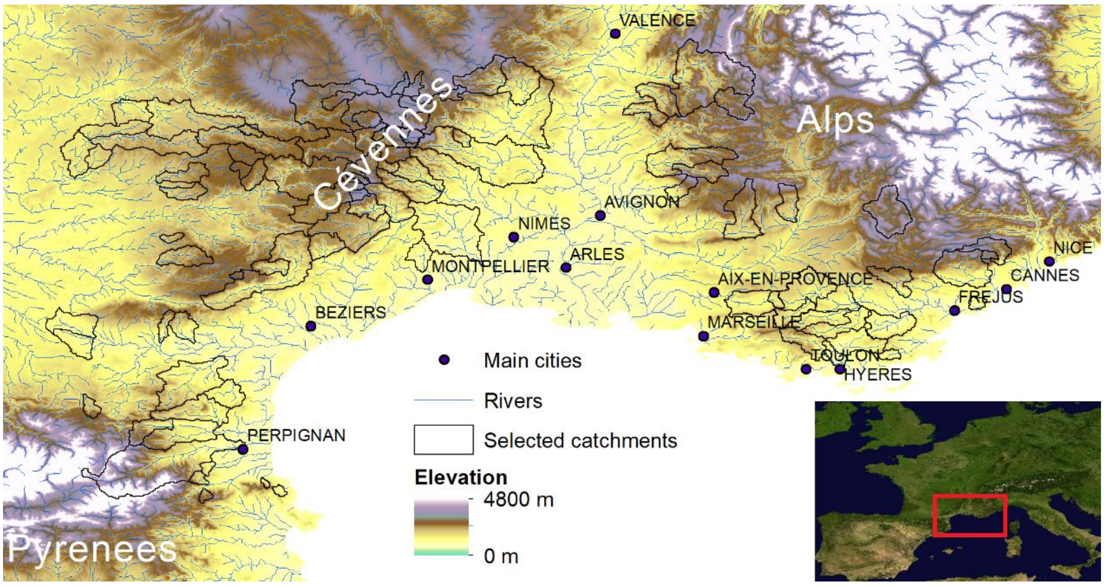
This is explained in the section 2. We excluded basins with a nival type of hygrographs (ie. snowmelt-driven peak discharge in spring) and with more than 20% of precipitation as snow. This is mainly since for French Mountainous basins, the SIM product may not be optimal source of data notably for extreme rainfall, see Gottardi et al 2012 and Blanchet et al. 2021. The focus here is on Mediterranean floods induced by rainfall.

Gottardi, F., Obled, C., Gailhard, J., & Paquet, E. (2012). Statistical reanalysis of precipitation fields based on ground network data and weather patterns: Application over French mountains. In Journal of Hydrology (Vols. 432–433, pp. 154–167). Elsevier BV. https://doi.org/10.1016/j.jhydrol.2012.02.014

Blanchet, J., Blanc, A., & Creutin, J.-D. (2021). Explaining recent trends in extreme precipitation in the Southwestern Alps by changes in atmospheric influences. In Weather and Climate Extremes (Vol. 33, p. 100356). Elsevier BV. https://doi.org/10.1016/j.wace.2021.100356

**Figure 1: It would be better to add an inset map to show the location of the study area in Europe.**

We provided a modified map:




**L155: I understand the reason for using POT1 instead of AM, but please clarify if this will affect the subsequent trend analysis.**

It does not have an impact on the results. The resulting distributions are almost identical, and actually numerous papers are using this type of sampling. See some recent ones:

https://doi.org/10.1038/s43247-021-00248-x

https://doi.org/10.1002/wat2.1520

https://doi.org/10.1016/j.jhydrol.2019.05.054

https://doi.org/10.1016/j.jhydrol.2021.126994

**L164: Please specify whether rainfall is the precipitation that excludes snowfall.**

No, we use total precipitation. But as mentioned above, the contribution of snow is small so it is not irrelevant to use the wording of "rainfall".

**L165: Is the precipitation on the same day as the flood peak considered, please clarify.**

Yes, we added it in the text.

**L174: How is the duration of the flood event calculated, i.e. how are the start and end points of the event determined? Also for the runoff coefficient calculation process.**

We slightly modified the description in the method section to: "Then, for each flood event, we computed the total rainfall and maximum rainfall. The *n*-day previous precipitation is extracted. Total rainfall for each event is estimated by a cumulative sum of precipitation starting the day of the flood and this aggregation stops if there are two consecutive days with precipitation close to zero (1 mm) to account for rainfall intermittency within events. The maximum daily precipitation is extracted from the same time interval used to compute total event precipitation."

For runoff coefficients we added: "The runoff coefficient was computed for each event as the ratio of direct runoff depth and total event precipitation".

**L183: How is September 1 of the hydrologic year determined?**

In the Mediterranean, the summer is dry. It is quite standard practice to consider that the hydrological year in this region starts in September.

**L197-206: As I noted earlier, the classification reads rather arbitrary.**

See our previous comment.

**L207: What does "other" mean?**

It is the name of the category of events that cannot be classified. As mentioned, this proportion remains very small.

**L220: What test was used to check statistical significance?**

As written the line above: the Mann-Kendall test for trends.

**Figure 2: what is the unit of "relative change"? I also suggest showing the statistical significance of these changes in the map.**

Relative change is unitless. If you multiply by 100 you get a percentage. We added in the caption: [-].

We choose to not add the information about local significance in the maps, with 6 sub-panels, 98 points per panel it would be impossible to see.

**Table 1: one-tailed or two-tailed test? please specify.**

Two-tailed. We added it in the method.

**L262-264: How is the conclusion related to the results?**

As shown in Figure 3, there is an increase in runoff coefficient with increased soil moisture.

**L269-273: I would expect a figure to support the results.**

We added a figure S1 showing the correlations in supplementary materials.

**Figure 3: please give the number (of events?) in each bin.**

We modified the figure to include the number of events:

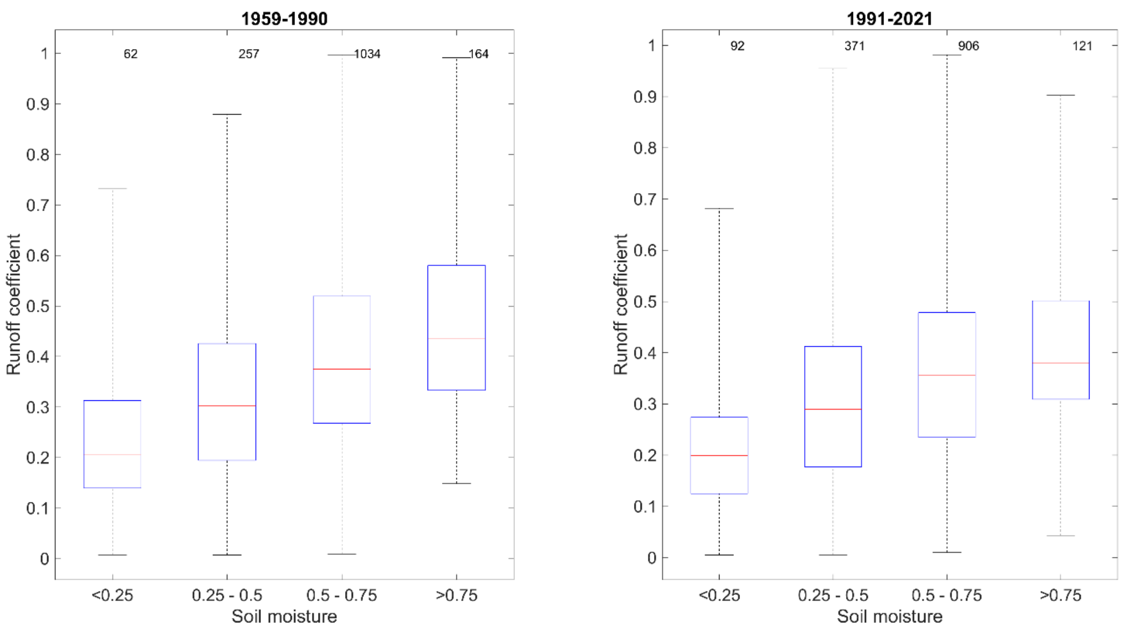


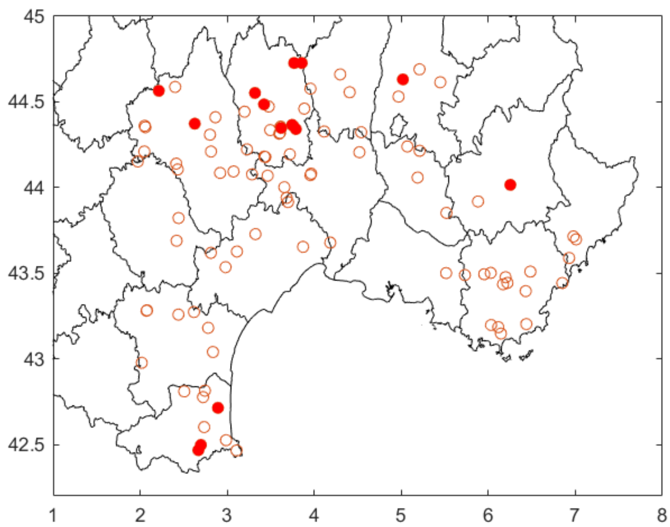

Figure 3: Relationship between the flood event runoff coefficients and antecedent soil moisture for the two time periods considered: 1959-1990 and 1991-2021. For each box, the central line indicates the median, and the bottom and top edges of the box indicate the 25th and 75th percentiles, respectively. The whiskers extend to the most extreme data points. The numbers at the top of the figure indicate the number of events in each category.

**Figure 4: not quite sure if POT1 events instead of AM events will introduce a bias in the flood dates.**

It is not clear whether this is a question or a comment. With POT we have floods in the sample, while with AM it is possible for some very dry years to have an annual maximum that is not really a flood event. Over all, best practice is to avoid introducing this type of events in the sample so POT is preferable (https://doi.org/10.1038/s43247-021-00248-x ).

**L298-302: it is difficult to follow without showing these basins on the map.**

In these lines, we mentioned a few cases where we observe a second minor peak of occurrence for floods, we added the following figure in supplementary materials:




Location of basins where a second minor peak flood is observed during spring (in red)

**Figure 5: Please also show the results of the significance test in the maps.**

Same comment as above, it would be difficult to see. As replied above, we added a regional significance test since the important question here is whether these changes are regionally significant or not.

**L304: Which time period (1991-2021 vs. 1959-1990) was referred to?**

We added: “1958-1990 and 1991-2021”

**L341: I would like to add significance tests on the difference between the proportions, which would be more supportive of the discussion.**

We applied the Chi-square test to assess whether the changes in the monthly frequency of the different weather types are different, but they are not significant (nor at the 5% or 10% level). Yet, as also replied to Reviewer 1, it should be reminded here that we are dealing with extreme events. So, a small number of events. We argue that even small changes in the frequency of flood-inducing weather types might have an impact on flood frequency. And beyond the "statistical significance", we don't know exactly how these changes in numbers may affect seasonal shifts in flood frequency. For example, in August, the frequency of WT4 increase from 8% to 11%, similarly in June from 11% to 14% and overall this represents an increase of +69 episodes (so 2 per year, that has to be related to the mean occurrence of floods in our study, one per year on average). We added this discussion in the text, also following recommendations of Reviewer n°2, first to clearly state that there is no "significant" changes, but also that more research on that particular topic is required =

"The seasonal patterns observed for the floods are closely related to the occurrence of different weather types in different sub-regions. As shown in figure 6, most basins located east of the Cévennes mountainous range have floods associated with WT4, Southern Circulation, and western basins with WT2, Atlantic circulation. The most frequent pattern associated with 37% of floods, WT4, is known to be triggering intense rainfall events in this region (Ducrocq et al., 2008; Tramblay et al., 2013). Interestingly, the WT6, Eastern circulation, and WT7, Southwestern circulation, are both associated to a lesser extent with floods across the whole region, but without notable spatial differences in the relative frequency of floods associated with these weather types. Change in flood seasonality, to certain extend, could be ascribed to changes in the seasonal occurrence of the weather types (Figure 7): WT4 tends to occur more frequently from March to August during 1991-2021 compared to 1959-1990. However, it should be noted that these change in the frequency of WT4 are not statistically significant according to the chi-square test and the relative increase is rather low, from +0.53% in May to +2.73% in August. When looking at the actual count of WT4 days, this change represents an increase of 69 events during that 6-month period for 1991-2021, so an average of +2.2 days per year. Associated with a warmer Mediterranean Sea over the last decades notably during summer (Pastor et al., 2020), the combination of these two factors could possibly explain the earlier occurrence of floods east of the Cévennes mountainous range. Similarly, there is an increased frequency of WT2 in January, February and March between 1991-2021 and 1959-1990. As for the changes in WT4, these increased frequency of WT2 in January to March is not statistically significant, and represents an increase of +65 days of WT2 during that period that could be possibly related to the later occurrence of floods west of the Cévennes range. Although this change in seasonality is a plausible cause of the observed changes in the flood seasonality, more research is needed to better understand these processes and attribute changes in flood seasonality. Notably, to analyze in more detail the moisture supply from the Mediterranean or Atlantic seas, the interaction with the atmospheric thermodynamics, the duration, localization and the spatial dependence of the rainfall episodes inducing floods."

**L398: I think we should be cautious about this conclusion, given the relatively high p-value and low correlation coefficient.**

We agree and changed the text to: **"**There is a significant, yet low, correlation ($\rho$ = 0.26, p-value = 0.008) between the ratio of excess rain floods and catchment size, with a larger proportion of excess rain in larger basins, while on the opposite there is an even weaker and negative correlation ($\rho$ = -0.16, p-value =0.09) between the ratio of short rain and basin size."

Yet, we would like to stress that we do not provide any strong conclusion here, just showing the results.

**L431-434: How statistically significant are these results?**

See one of the first comment above. Catchment-by-catchment it would be difficult, if not irrelevant, to compute the 'statistical significance' given the small sample size. See the results of Figure 11 showing the regional changes, showing that regional changes in the frequency of excess rain and short rain floods are significant. We assess the regional significance in section 4.5.

**Figure 10: Please also include the significant test results. I am also not clear how the change in frequency is calculated for each catchment. Do you compare the frequency in 1991-2021 with the frequency in 1959-1990?**

Yes, we compare the frequency of the different flood types in the two periods. The regional significance of these changes is shown in figure 11. Again, the scope of the study is to assess whether these changes are significant regionally and not necessarily catchment-by-catchment. Given the small numbers, notably of Short rains in some basins, the regional approach is much more relevant.

**L445: How was the conclusion reached? I don't understand the logic. I am not surprised that the trend level is not consistent. Even a small change in the driver magnitude can lead to a change in the flood type (because the threshold leads to a hard boundary in the classification, that's why I asked for a sensitivity analysis).**

As shown above, the results are not strongly dependent on these thresholds, that are actually quite standards to define precipitation extremes or soil moisture state. What we show here, is that trend magnitudes can differ from one catchment to another, given different catchment sizes, land use, the presence/absence of groundwater contribution, so this result seems absolutely expected. As you requested, we included a sensitivity analysis and the main conclusions are not affected by threshold selections.

**Figure 11: Please show the 25th and 75th percentiles of the regional frequency to show the spatial variance.**

On figure 11 are plotted the annual observed frequency of the different flood types. So, it relies on counting for each year the number of Excess rain floods, Short rain floods and Long rain floods, in all basins, and this count is divided by the total number of events each year. It is not the mean frequency over all basins here, so the computation of $25^{th}$ and $75^{th}$ percentiles is not applicable.

**L477: "This is mainly due to a decrease in the specific discharge of short rain floods" I can't understand it, because if the short rain has been observed intensified, if we only consider short rain floods, it is more reasonable that the flood magnitude will also increase.**

No, the short rain floods are not intensifying, They are just more frequent.

See Figure 12. On average floods induced by short rains have a larger specific discharge than other types of floods. Over time, there is an increase in the number of floods induced by short-rains. But, in the same time, the magnitude (or severity) of this type of floods does show a slight decrease, and not an increase, in terms of specific discharge.

We rephrased this section to =

“Given that there are different flood sample sizes in the different basins corresponding to different flood-generating processes, we pooled regionally the flood events. To do so, we computed the specific discharge for each event (i.e. the flood magnitude divided by catchment area) to analyze the distributions of specific discharge for all the events associated with excess rain, long rain or short rain. Specific discharge is used herein since it is a good indicator of flash floods severity, notably in this Mediterranean region (Delrieu et al., 2005, Ruin et al., 2008). Figure 12 shows that the short rain floods are more severe, in terms of specific discharge, than excess rain or long rain floods at the regional level (as shown also by Tarasova et al., 2023). The regional distributions are different according to the Kolmogorov-Smirnov test. It must be noted that for a given basin the magnitude of the different types of floods may not be very different, showing the strong variability from one event to another that is not solely linked to the flood trigger. When comparing the different flood distributions between the time periods 1959-1990 and 1991-2021, the differences in flood magnitudes between excess rain, long and short rain are reduced. This is mainly due to a slight decrease in the specific discharge of short rain floods, notably for flood events with a return level higher than 10 years, while the excess rain floods show very little changes in intensity over time. “