NB: also not required by the Reviewers, we propose to change the title into "A statistical framework to link the torrential event occurrence to regional and local driving atmospheric conditions – the case of the Northern French Alps", in order to 1) better highlight the fact we conduct a statistical analysis rather than an event-based analysis ("A statistical framework"), 2) better highlight the generality of the method that could be applied elsewhere ("the case of the Northern French Alps"), 3) better stress that we only use dates of events and no other hydrometeorological data ("torrential events occurrence").

Answer to Reviewer #1

The current work submitted for review attempts to characterise and identify the critical atmospheric driving conditions of torrential (flash) floods at a local scale using an event dataset collected over the Grenoble Metropolitan area in France and then tries to extrapolate the same over a larger regional scale (Northern French Alps). While the methods applied seem robust enough and the results plausible, I have some significant concerns which should be considered before the paper can be considered for publication in HESS.

My main concern is related to whether the paper is the best fit for the overall scope and target audience of HESS. In my opinion, the current work does not significantly advance the current state of the art in research linked to such extremes and has some critical shortcomings.

=> We thank the Reviewer for his/her thorough review. We regret that the Reviewer does not find that the article significantly advances the current state of research. However, as noted in the introduction, we are aware of only one statistical analysis on the atmospheric conditions generating **torrential** floods - Turkington et al. 2014 - , whereas many works (including those of the co-authors) have been done on the atmospheric conditions generating **riverine** floods. We would be grateful if the reviewer could provide us with further studies on the subject of atmospheric conditions for torrential events, as we indeed claim that the advance of our work is mainly due to the fact that there are gaps on this subject.

Critical Comment

- In the beginning of Methods (203-204): the authors remark that "*Our goal is to determine which atmospheric variables are very different from the climatology the days of the torrential event*". However, in the 70 events considered for the subsequent analysis (117-118), *b) one or several torrents and no river (purely torrential events), c) one or several torrents and one or several rivers (mixed events)* are considered. Maybe I missed something…wouldn't this lead to a situation where flooding episodes in which riverine processes (saturation excess/snowmelt runoff) played a major role are also included in the torrent event classification? The authors seem to neglect the purely riverine events. In my opinion, the riverine events could have been utilised to overcome the limitations of including the mixed events, or I would have only used purely torrential events.

=> Please let us recall that the considered events are extreme – the torrential events correspond to return periods of order 2-3 years at the scale of the conurbation and of 15-170 years at the scale of the torrential units. We don't think it is possible de get such extreme torrential events due to snowmelt or saturation processes only. Also please note also that mixed events correspond to dates where we had both riverine and torrential events, however the events did not necessarily occur at the confluence between torrents and rivers (anyway, as shown in Fig 1 of the article, many torrential units have no confluence with rivers). So it seemed reasonable to us to include the mixed events in the analysis because these are dates where something did happen on torrents. Of course, it would also be interesting to understand why some dates are only torrential whereas other are mixed, however we unfortunately don't have for that enough data to conduct a statistical analysis as in this article. Finally, we don't think that including purely riverine events would be helpful because then we would potentially mixed atmospheric conditions triggering flash (torrential) flood over a few hours and atmospheric conditions triggering riverine flood over several days (the time of concentration of the Isère in Grenoble is about 3 days) – these are probably very different conditions. Finally, let us recall that many works have been done on the atmospheric conditions triggering riverine floods, in particular in this region (see e.g. Blanc et al. 2022), whereas works on torrential floods are missing. This is the reason why we wish to focus in this article on floods only.

- The authors state that Comparison of 20CR-2 and ERA5-3 *allows assessing whether different reanalyses see the same driving atmospheric variables over the recent period and the same region*. However, they then proceed to *Furthermore events in the HP class are quite discordant between 20CR and ERA5 - half of the HP events of 20CR over 1950-2014 are otherwise classified in ERA5. For these reasons, the HP class is excluded from the analysis.* Isn't it also worthwhile to check why the coarser scale 20CR and finer ERA5 has difference in the HP class (which from Fig 2 looks to be the only class having a high pressure system over the study area). It is at least expected that the authors provide more explanation on the discrepancy before just ignoring them in subsequent analysis. They also exclude the E-NE class stating it has very few events and is not usually associated with high precipitation; however, that only makes it more worthwhile to explore why such events occur and what are the driving unusual atmospheric conditions behind them.

=> We understand the need for a deeper explanation with HP classes differ between the two reanalyses. However please note that "half of the events" means actually "2 events over 4", so the primary reason why we don't include the HP class is in the previous sentence: the HP class contains too few events to conduct a statistical analysis ("very few events are classified in the HP class after 1950 with either database, giving a very low probability of sequences of the HP class to generate torrential events"). We will investigate the difference for these 2 events but, as already showed in Blanc et al 2022, 20CR fields are smoother and more regular than ERA5 fields. This explains why it classifies too many days in HP, as also shown in Table 2 of the article (37% of 3-day sequences in the HP class in 20CR versus 28% in ERA5). So there is probably nothing more to these two events than the fact that the HP class is overrepresented in 20CR.

We also understand the suggestion of the Reviewer to investigate what are the unusal conditions driving E-NE events, however in this class we only have few events (4-5 events), thus it is again impossible to conduct a statistical analysis as in the rest of the article. The only thing that could be done is to look at the atmospheric variables for the few E-NE events, one by one, and comment on what we see but we fear it would unbalance the article as it would be an "event-based analysis" rather than a "statical analysis" as for the rest of the article. Finally, please note, as stated in the article, that the N-W, SE-SW and BS classes contain 75 to 87% of the events depending on the reanalysis, so by excluding the HP and E-NE classes, we only exclude a minority of the events.

We will clarify these two points in the new version of the article.

- I am also a bit confused about the spatial scales and the averaging used in this study. Are the NEP calculated for each grid point for both 20CR and ERA5 or do you average out all the grids and then calculate the probabilities based on the spatial average time series? Maybe I missed this information; however, this should be emphasised together with the uncertainties involved.

=> We apologize for the confusion. We apply your second alternative: we average out the atmospheric variables over the grids points. This gives us, for every atmospheric variable, a 1-dimensional daily time series. Then we compute the NEP associated to this time-series. We will make it clearer in Section 2.2.3.

- The paper lacks a good discussion section linking the results of the work to the overall research gaps in the field of both extreme events and hydrology. While interdisciplinary studies like this are indeed valuable for the hydrological community, it is expected that the authors will provide a strong discussion linking their atmospheric research to catchment scale and more local water cycle changes.

=> Please note that we have no other data on the torrents than the event dates – this is by the way the reason why we built the database of dates based on the RTM reports, because we have nothing else! There are in the area no series of torrent discharge, only one long-term rainfall station, we don't know precisely the geomorphological characteristics of the torrent catchments, and of course there is no calibrated hydrological model. So unfortunately we are not able to link our results to catchment scales or local water cycle: the only thing we have is a series of dates. This is a very important point that we did not stress enough in the article. In particular, it'd be impossible for us to conduct an analysis such as that of Prenner et al. 2019 that relies on many catchment data. Our analysis can be seen as a "low-cost" analysis where only dates are necessary to link them with atmospheric conditions. We will clarify this important point in the Data section of the article, as well as in the last part of the introduction that will be partly rewritten. We also propose to mention "torrential event occurrence" in the title to make clearer that we use dates only.

However, we understand the point of view of the Reviewer that the paper lacks a clear discussion. Actually, in the current version the results and discussion are mixed in Section 4, which we understand is confusing. Thus we propose to split Section 4 in 2 parts: a "Results" Section that would only be about the NEP values (including Figs 4, 5, 9), and a "Discussion" Section that would comment on the maps of the discriminant variables of each class (including Figs 6, 7, 8) and discuss how they differ from what we know of riverine flooding in the literature. We hope this will clarify the article.

I also had a feeling that some of the content of the manuscript seems very clear and straightforward to the authors but is not so clear to more general readers. For example, Figure 1 doesn't follow the acceptable standards for publication in an international journal. The north arrow and scale bar are missing. It is not clear to me what a torrential unit means. Is it linked to the geomorphology of the study region, or is it linked to the actual event reporting convention followed by the agency?

=> You're right, we will add in Figure 1 the north arrow and a scale bar and go through the article to see if any other information is missing. The torrential units are a definition of the agency.

- In the conclusions, the authors noted that – "*A remarkable result of our study….. the discriminant atmospheric variables are as discriminant whether we consider them at local or regional*". Are these results really remarkable? The authors themselves note that (185-190), as

Grenoble is at the border between 2 pixels of 20CR, we cannot consider a smaller region than the whole French Alps for 20CR. Doesn't this mean we move ahead with the assumption that local-scale site specific atmospheric conditions are indeed related to the dynamics captured by the coarser scale product? I would recommend defining the local and regional scale more strictly here to close these inconsistencies. Furthermore, spatially averaging (which is not well defined in this paper) at the so-called local and regional scales would add significant uncertainties to this methodology. By default, the spatial averages at both scales should indeed contain information complimentary to each other.

=> Actually the "remarkable result" relates to the comparison of ERA5-Alps and ERA5-Grenoble. In the first case, we average the ERA5 atmospheric variables over the 105 pixels of the French Alps before computing the NEPs. In the latter case, we average the ERA5 atmospheric variables over 9 pixels covering Grenoble conurbation before computing the NEP values. Remarkably, in both cases, we find the same discriminant atmospheric variables and with very similar Silhouette values (see Fig 5, yellow vs. red curves). This means that, although we consider date of very local events (the torrential units scale a few km^2), we find the same "abnormal" atmospheric variables whether we look at them at the scale of the French Alps or at the scale of Grenoble. This seemed remarkable to us and actually we did not expect that (we expected less discriminant NEP values at the French Alps scale). If the word "remarkable" causes concern, we can replace it with "unexpected".

- The authors end the Introduction by stating – "*Our work makes three contributions to the work of Turkington et al. (2014).*" I expected a much more concise presentation about the work and its proposed relevant outcomes for understanding hydrological functioning rather than simply stating how this work is different from another similar study carried out around nine years ago. Furthermore, the first advancement is confusing *(As far as we know, no study applied such an approach at torrential scale before the 1950s*), as it is later stated that the period before 1950 is not being investigated fully due to inconsistencies in the 20CR model (*Given these discrepancies, we were unfortunately unable to study the nonstationarity of the driving atmospheric conditions. Thus, the rest of the paper focuses on the post-1950 period*)

=> We understand your point of view and this part (last paragraph of the Introduction) will be deeply rewritten.

Minor comments:

- In conclusion, the authors again stater *The fact that we were able to find discriminant signatures at regional scale is a good hope that global climate models such as CMIP6.* To my understanding, the grid resolution of CMIP6 is at least comparable to the resolution of the 20CR model used here. Why not use the historical CMIP6 model ensemble directly at the so-called regional scale?

=> Actually historical CMIP6 simulation do not reproduce the true climatology (they are simulations) so we could not link them to actual dates of events. Anyhow, ERA5 is of better resolution and it's the only database that allows to compare two spatial scales (French Alps vs Grenoble).

- The abstract needs to be restructured for better readability. I would omit the part about applying the method elsewhere..

=> We will work on the abstract to make it more readable and more concise.

- The terms ERA5-3 and ERA5-4 is actually confusing because you also talk about 3 day windows.. I would recommend naming them as ERA5-A for Alps and ERA5-G for Grenoble.

=> Thanks, this is a good suggestion.

- *We note that the two periods have different lengths (99 versus 65 years), however considering two equal periods almost does not change the results due to the absence of events in the 1930-1940s* – It is not clear to me what is meant here.

=> We meant here that, since we're only interested in the NEP values of the torrential dates, considering 20CR on 1851-1949 (99 yrs) vs 1950-2014 (65 yrs), or on 1851-1932 (82 yrs) vs 1933-2014 (82 yrs) would barely change the number of points in e.g. Fig 4 since there is almost the same number of events in 1851-1949 and 1851-1932. But anyway, in order to simplify the article and to make it less confusing, we will remove the analysis of 20CR prior to 1950 since anyway we conclude that the quality of 20CR in the remote past is less good, so this part will disappear. Thus we will in the next version focus on three databases: 20CR-Alps on 1950-2014, ERA5-Alps on 1950-2019, ERA5-Grenoble on 1950-2019. Another possibility to make the article shorter could actually be to keep ERA5 only (ERA5-Alps on 1950-2019, ERA5-Grenoble on 1950-2019) since anyway 20CR gives the same results as ERA5-Alps.

- In Figures 6-8, the authors only show one atmospheric variable (IVT, CAPE & PWAT) for each class (NW, SE-SW and BS respectively), however, it would be interesting to see the three variables for all the three events so that a ready comparison is possible. This could help strengthen the argument about different drivers for different events.

=> We understand you're point of view. We will show in the next version other variables than the discriminant ones but note that we may have to make a selection since there are 7 variables in total and showing all will make big figures.

- I could also find quite some missing words and grammatical inconsistencies. The draft could benefit from a fresh reading to smoothen out such errors.

=> We will carefully go through the article to check for typos and errors.

Overall, in view of the critical shortcomings, I recommend the paper to be rejected.

=> We sincerely respect your recommendation, and we thank you again for your comprehensive review. However it seems to us that the above comments could be addressed in the next version of the article. I hope you will find our replies convincing.

References

Blanc, A.; Blanchet, J. & Creutin, J.-D.
Characterizing large-scale circulations driving extreme precipitation in the Northern French Alps
International Journal of Climatology, **2022**, *42*, 465-480

Prenner, D.; Hrachowitz, M. & Kaitna, R.
Trigger characteristics of torrential flows from high to low alpine regions in Austria
Science of The Total Environment, **2019**, *658*, 958-972