Reviewer #2

Summary

The paper analyses recent daily and hourly precipitation extremes in the French Alps using in-situ station data from Météo-France and Électricité de France. Looking at 20 year return levels, the authors present an updated climatology of precipitation extremes as well as a trend analysis. The connection to dominant weather patterns during extreme events provides hints as to the drivers of the observed extremes. The analysis is valuable not only for stakeholders in the region, but to the broader scientific community in providing insights into changes of precipitation extremes in topographically complex regions and the drivers thereof. This is especially useful since observations of extreme precipitation are generally short and model projections are yet highly uncertain. I recommend a publication of the manuscript after the following comments are addressed.

Thank you for your detailed comments! We greatly appreciate the time and effort you put into reviewing our paper and the constructive feedback. Please find our answers below in blue.

General comments

Abstract: You focus largely on the motivation and the methods in your abstract. You only describe the results on a very high level and do not mention any interpretation, implications or needs for further investigation you identified. Please extend on this.

Thank you for this comment, we will reformulate the abstract accordingly.

Study area and data: Your presentation of the station data is very confusing:

1. (1) I count 32 EDF old data stations in Fig. 1, but the text in Section 2.2.2 states there should be 27 stations for the long term data, circled in Fig. 1. It is also unclear to me how these 27 are exactly derived from the 65 rain gauges, only 48 of which are from EDF (are only these 48 considered or all 65?), that have long term data, and the 68 provided by EDF with more recent data. You state that these are derived by “combining old and recent data” but it is not clear to me what that means. Do some or all of the 68 recent stations overlap with the old 65, or 48? There seem to be 68 EDF stations (old and recent together) in Fig. 1, so it seems to me that the old datasets always also have recent data because they always have colored circles. You state that if “the two datasets were overlapping, the more recent dataset was used”. Does that mean that there are two separate stations in the same location?
2. (2) I only find 83 Météo-France stations in Fig. 1 as opposed to 89 mentioned in 2.2.1.
3. (3) The introduction and Section 2.2.5 state that there are overall 177 time series in the final "recent" dataset. Summing up the (potentially) 89 time-series from Météo-France with the 68 recent EDF stations only yields a total of 157 time-series. Please elaborate where the other 20 come from. I only count 151 stations in total in Fig. 1, and about 160

the “Annual” panel of Fig. 8. Please check all of your figures and numbers in the text for consistency. If any station is for some reason excluded from any analysis, either exclude it from the whole dataset or clearly describe the procedure which stations are included were.

4. (4) In section 2.2.5 you furthermore state that you use 67 of the 177 for trend analysis. Please mark these either in Fig. 1 or reference Fig. 4 here. Fig. 4 only contains 66 stations according to my (admittedly limited) counting capabilities, not 67 as stated in 2.2.5, please double check.

Overall, you should considerably improve sections 2.1, and 2.2, as well as Fig. 1 to be as concise as possible about which stations you actually analyse and where they come from. Please integrate section 2.2.5 into a new overarching “station observations” section together with 2.2.1 and 2.2.2 or at least move it further up to be directly after the first two subsections on the observation data to avoid further confusion. Please also consider dropping unnecessary information in these sections as for example naming all individual départements in the text, the detailed description of the study area geography including individual valley and mountain names, as well as the history of the station networks in sections 2.2.1 and 2.2.2 (and the introduction). Lastly, please elaborate on the time period you are looking at, do the Meteo France observations also stop in 2022 as do the EDF data? From when to when does your 30 year period reach for the trend analysis?

Thank you for the detailed observations. We see that the presentation of the data is both unclear and inconsistent. We will therefore completely reformulate this part following your suggestions, and ensure that the numbers are consistent.
Concerning the details on the study area, we think they contain valuable information to many readers and would thus like to keep them. Also, we received positive comments on this part from reviewer #1. Keeping it in a separate paragraph should make it easy for readers uninterested in this particular part to skip it.

Neighborhood Approach and trend estimation:

1. (1) In section 3.2.1 you mention that you discard stations where the parameter estimation of ξ is unrealistic, please state how many these are and how many are then left in the dataset. Is this different from month to month / season to season?

   We did this in a previous stage of the project, but noticed that the problem could be resolved by constraining $-\infty<\xi<1$. It is our mistake that the sentence is still there, and we will remove it in the revised version of the manuscript.

   Please elaborate if the fact that your study region is in a highly complex terrain affects the validity of the neighborhood approach. Do you employ the neighborhood approach across the climatological divide? How robust is your assumption that stations close in space share a similar climate in this terrain, considering the dominant weather patterns causing extremes vary, even between stations that are close together?

We do indeed apply the neighbourhood approach with a fixed radius regardless of the topography. As we mention in l. 408 we tested for different radii and found that the spatial patterns remain overall similar. In our experience, the strong influence of orography on mean precipitation is much smaller or even non-significant for extremes. In previous studies (Evin et al, 2016, https://doi.org/10.1016/j.jhydrol.2016.08.024) we have used a 'crossing distance', taking orography into account, but during preliminary analysis for the study of Blanchet et al (2021, https://doi.org/10.1016/j.wace.2021.100356) we did not find that this was significantly better than the simpler Euclidean distance.
As we show in Fig. 8 and as also Blanchet et al (2021) show, the location of the climatological divide, with regards to extremes, is not fixed between months and also over time, so taking it into account explicitly does not seem relevant. It should also be noted that the climatological divide of Auer et al (2007) has been determined on the basis of several climatological variables and on the much larger scale of the European Alps. In our study, we thus take it more as a guide than a hard border.

2. (2) What exactly is it that you call "trend"? Is it μ1? If so, is the trend for stations that choose Mσ as their best model, automatically 0? Or do you fit a trend slope to the 20-year return level values inferred from the non-stationary GEVs? How is this trend slope calculated? Are the trends that you show then by definition from 1985 to 2022?
   A trend is the slope of the 20-year return level derived from the non-stationary GEVs (slope of the solid red line in Fig. 3A). These trends are indeed calculated over the period 1985-2022, and scaled to mm/d/10y for better interpretability. We will clarify this in the revised version of our manuscript.

Map figures:

1. Please plot the in-situ maps using a discrete colorbar to enhance the interpretability of these figures.
   Thank you for the suggestion, we will do that.
2. I find that the monthly maps do not add significant value to your interpretation and recommend moving them to the appendix. Most arguments you make can be made based on the boxplots or with a few references to the appendix.
   Thank you for this suggestion. We understand your point, however, in our opinion a map is more visual and thus easier to interpret with regards to spatial patterns, especially spatial coherence of trends. This is even more important as the location of the climatological divide varies between months and seasons, and thus splitting the region into north and south along a fixed line is to some extent arbitrary (see our reply to one of your previous comments).
3. The motivation for shifting the seasons for the trend analysis is well made and the maps in support of it can also be moved to the appendix.
   Thank you for this remark. In the revised manuscript, we will use only the 'new' seasons (JFM etc.) for both stationary and non-stationary analysis, and explain this decision in the relevant paragraphs of the Methods and Discussion sections. This will reduce the number of subfigures in Figs. 10 and 12 and improve the overall clarity of the

manuscript, while maintaining a coherence between the stationary and non-stationary analyses.

In this way, you can limit the panels in the map figures to always four seasons and the annual map, greatly reducing the overall number of panels in the main text and consequently improving on overall readability, keeping the readers focus on the significant findings.
For the reasons named above we would like to keep the monthly panels in the figures, but we will remove the traditional seasons in Figs. 10 and 12 (and change the seasons from traditional to new seasons in Figs. 5,8,9). Following the suggestions by Reviewer #1 we will additionally remove the right panel of Fig. 2, and move Fig. 4 to the appendix. We are also considering merging Fig. 5 and 8, coding the three WPs in Fig. 8 into symbols in Fig. 5. This should help reduce the overall number of figures and improve readability.

Please discuss the significance of the trends you find in more depth. Are there differences in significance between seasons/months or the daily and hourly aggregation levels? Do you have hypotheses as to why? How many stations actually show significant trends? You state in the abstract that trends in hourly extremes are rarely significant but do not mention this in the results.
We will expand the discussion on the significance of trends in our revised manuscript.

Specific comments

Line 8: "In the first part, we use a stationary framework....", you do not mention a second part in the abstract. Please do so and describe where you use the non-stationary framework and why.
We will reformulate the abstract following your first comment and take this into account.

Line 63: "Furthermore, it is expected that the trends in intensity should be higher for subhourly durations" please quickly explain why. Later in this paragraph you give explanations for super-CC scaling, please clarify if you therein focus on hourly or sub-hourly extremes.
We will restructure the paragraph to make this clearer.

Figure 1: This figure shows the study area together with rivers, départements borders, selected cities, orography and the three different station types. Please omit numbers and names of départements that are not actually used in the study. Moreover, please plot the stations above river names or département numbers, some of the stations are almost completely (and some maybe completely?) covered by them, see for example the “Drac” river label.
Thank you for these observations, we agree that labels should not cover stations, and will modify the figure accordingly. Please note however that some few stations are very close to each other (less than a few hundred meters apart), and will thus not appear separately on the map (for example in Villard-de-Lans, southwest of Grenoble).

Line 131 and 136: Please explain what "structured" means in this context.
Orographically structured. We will clarify this.

Lines 139 to 155: Please provide some references for this climatological assessment of the study region.
Thank you for pointing this out, we will do that.

Line 157: You already mention EDF as Électricité de France in the Introduction, please give the abbreviation at the first mention in the typical format " Électricité de France (EDF)".
Ok.

Line 175: "a change from mechanical to digital registration of observations took place in the 1980s, which seems plausible given the obtained dataset": please explain how this "seems plausible".
The ‘old dataset’, of which we only have access to weekly maxima, ends sometime in the 1980s (different end date between stations). We assume that this coincides with the change to digital registration. We will clarify this in the revised version.

Line 184-189: The description of the quality control procedure should go into its own subsection, for the sake of not distracting the reader here. You could for example rephrase the part here to “We employ a quality control procedure described in section 2.2.x. One station was removed during this procedure, yielding now a total of ??? stations in the dataset.” Moreover, since we do not know the names or numbers of any of the other stations, and neither the quality-controlled nor the raw data is publicly available, it is not very meaningful to provide this for the one station that is removed, in the text.
Thank you for this suggestion. Following your earlier comments we will completely restructure the data section, and take this into account.

Line 192: Please explain what you mean with the higher resolution of the recent dataset. Do you mean higher temporal resolution?
We do indeed mean higher temporal resolution. We will clarify this.

Caption Fig. 2: Please provide a reference for the LOESS smoother.
Following reviewer #1, we will remove the right side of the figure altogether. Nevertheless thank you for pointing this out.

Line 226: "the maximum of year and season or month m" please rephrase for improved understanding, e.g. "the maximum of season or month m in every respective year t"
Ok.

Line 159: Please give examples of the mentioned literature where ξ is kept constant.
We will add this.

Line 264: 100 resamples don’t seem very many?
We will increase the number of resamples to 500.

Line 273: Please briefly elaborate on the theoretical considerations that render the bounds $-\infty < \xi \leq 1$ reasonable or give a suitable reference.
Ok.

Fig. 4: Please also plot the climatological divide here and mention in the caption that these are the stations considered for the trend estimation.
Following Reviewer #1, we will move this figure to the appendix, but we will add the climatological divide.

Line 337: "As the weather pattern classification by Garavaglia et al. (2010) used here is done at daily scale, it is less relevant at hourly scale so we choose to not show the detailed maps here." Please clarify that while the reference work only uses the technique at daily scale, you still apply it to the hourly scale and why you conclude that it is less relevant. Just because the reference did not do so does not mean it is not relevant. In my opinion, the fact that Anticyclonic conditions are more dominant drivers for hourly extremes in summer, is a hint for a larger role of convection for these extremes and is, in fact, very relevant. Please provide the maps for this analysis at least in the Appendix.
Thank you for this comment. Applying these weather patterns for hourly precipitation is less robust as a consequence of the methodology used by Garavaglia et al. (2010) to classify the different weather patterns. However, we agree with you that it still bears some relevance.We will also replace the word 'relevant' by 'robust'.
Reviewer #1 suggested to merge Figs. 5 and 8 (daily stationary return levels and weather patterns). We will try this, and if the resulting figure remains readable we will do it for both daily and hourly stationary 20-year return levels. If it does not give a readable figure, we prefer to show the maps showing the weather patterns generating hourly maxima in the appendix in an interest to not add additional figures to the main body of the manuscript.

Caption Fig. 8: "generating monthly, seasonal or annual maxima..." Please add. Also please make sure all figure captions are harmonized (e.g. "First and second row...." description for all or none of them).
Thank you, we will do that.

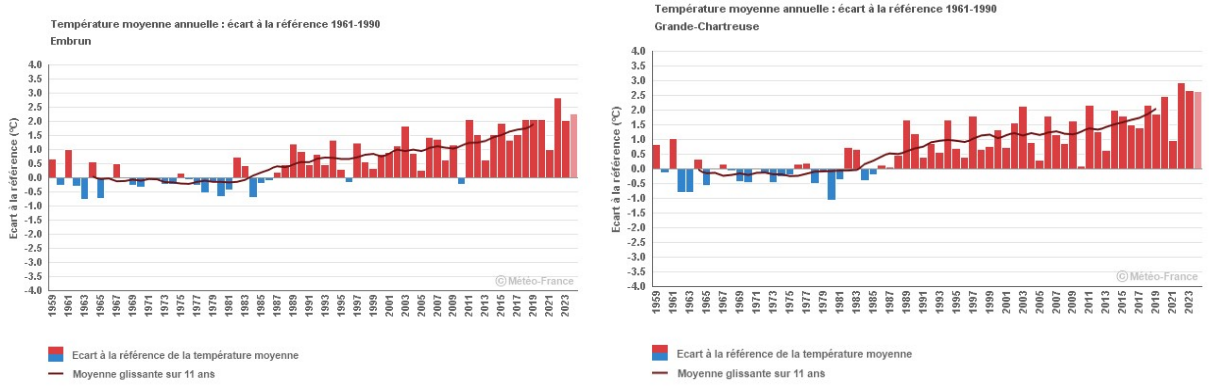
Line 346 and later: It is sufficient to include references for the climatological divide the first time you describe it and can be omitted thereafter.
Ok.

Line 351: "They show significant variability temporally (between months and seasons)" → they show significant variability between individual months/seasons - to avoid confusion. Similarly, please change in lines 363-264.
Ok.

Figures 6 and 11: If you cut off the outliers, one would be better able to see the patterns in the

boxplots, especially for the trends in hourly return levels, where the boxplots become very small. In any case, please state at which quantile the whiskers are cut off / what is considered an outlier.
Thank you, we will do this for the hourly values.

Line 394 ff.: When you contemplate on the suitability of time as the sole predictor of your non-stationary return level estimation it would be useful to provide evidence that temperature in the study region increases somewhat linearly with time, motivating that a climate change signal also in extreme precipitation should reasonably correlate with time.
Thank you for this remark. Météo-France have done this for selected homogenised time series of temperature measurements, the results of which are available at https://meteofrance.com/climathd. For example, the stations of Grande-Chartreuse and Embrun are located in our study region and show approximately linear temperature trends since the mid-1980s (Figs. below, see also Gibelin et al., 2014, DOI:10.4267/2042/54336). We will mention these in the revised version of the manuscript as well. Actually, these figures provide another reason as to why choosing a breakpoint with linear trends starting in 1985 as we do in our study seems reasonable.
It should also be noted that the best choice of covariates is an ongoing discussion in the scientific literature (e.g. Roderick et al, 2020, DOI:10.1029/2019WR02692, or Schröer & Kirchengast, 2018, DOI:10.1007/s00382-017-3857-9), and entering this discussion would be out of the scope of our paper.



Mean annual temperature anomalies for Embrun (Hautes-Alpes) and Grand-Chartreuse (Isère). Figures taken from https://meteofrance.com/climathd with data by Gibelin et al. (2014, DOI:10.4267/2042/54336)

Line 406: "However, in our experience this is difficult to implement and not robust, which is why we decided against using this method in our study." Please elaborate on how the mentioned approach is not robust.
In our experience homogeneity tests such as the Hosking and Wallis test tend to be too permissive, i.e. they tend to falsely classify nonhomogeneous region as homogeneous thus resulting in too large neighborhoods (see Evin et al., 2016 for a simulation study). We will clarify this in the manuscript.

Line 415: "we here arrive at the limits of studying precipitation extremes based on surface observations." Please add "to date" at the end or similar, to express that you expect future studies to perform better, not that a trend estimation in extreme precipitation is never and will never be possible with surface observations.
We will do this, thank you for pointing it out.

Line 444 f.: "Second, the spatial distribution of hourly extremes, showing higher return levels along the southern and western slopes and pre-Alps compared to the inner Alps, a pattern consistent with the findings of Fukutome et al. (2015); Panziera et al. (2018) for Switzerland." Please share with the reader hypotheses as to why these patterns are observed if available.
When unstably stratified air masses reach the first mountain ranges, convection can be triggered by the orography. Outside the mountains as well as in the inner Alps, convection occurs mostly due to spontaneous thermodynamic triggering. Fukutome et al. (2015) also note that there is a lack of sufficient moisture convergence in the inner-Alpine zones. We will add this to the manuscript.

Line 448 ff.: You state that you find "clear differences between the north and the south of the region" for both daily and hourly trends in extreme precipitation. However, in section 4.2.2 you state that "the clear north-south pattern in trends found for daily extremes in autumn is completely absent in the hourly data" and do not mention a clear north south divide for the other seasons; rather you talk about a potential east-west gradient. Please streamline the statements made to be better aligned, presenting a potentially more differentiated view on the data in both sections.
Thank you for highlighting this, we will reformulate the relevant parts accordingly.

Line 468: "but also not totally in line with..." please rephrase, colloquial. It is not clear how the hourly trends being different from the daily ones would support consistency with CC scaling and how both being similar does not. Moreover, it would be useful to give the reader information about the warming in this region over the study period. Without this, it is hard to infer if the results are in line with CC or not.
We acknowledge that we have to reformulate this part. We will include the references to ClimatHD (https://meteofrance.com/climathd) and Gibelin et al. (2014) (see our reply to your previous comment) showing positive and roughly linear temperature trends at selected sites in the study area. In light of this warming, positive trends in extremes would be expected as a consequence of CC-scaling. Negative trends, as we see here for daily extremes, or trends close to zero thus need to have a different explanation than the purely thermodynamic one, and in this section we provide possible reasons as to why this is the case.

Line 469: "Thus, these trends might be the result of…" → It is not clear to me how this statement derives from the finding before, consider omitting "thus" or improve on the explanation.
We will reformulate this section and take your comment into account.

Line 474 ff.: Please elaborate on the mentioned albedo feedback loop and how it would affect

rainfall, this is very interesting.
We will reformulate this. The idea is actually quite simple: surface heating due to reduced snow cover leads to increased thermodynamic instabilities, thus favouring convection (as shown by Giorgi et al., 2016, in a modelling study). For the French Alps, Beaumet et al. (2021, DOI:10.1007/s10113-021-01830-x) show particularly high warming trends especially in intermediate altitudes (1500-2000m a.s.l) in spring and high altitudes in summer, which they associate to an earlier melt of the snow cover at these altitudes.

Line 490 f.: Here you give trend estimates in %. It would be very useful to have changes in percent or percent per decade also in the Appendix. Otherwise, please put (not shown) here.
We agree, but given that we already promise to put a number of figures in the appendix, we prefer not to add this one as well (and therefore add 'not shown').

Line 505 f.: "Regarding recent trends in extremes, we find a mixed picture of both positive and negative trends in different areas and seasons." → "Regarding recent trends in extremes, we find a mixed picture of both positive and negative trends in different areas and seasons with considerable differences in patterns between daily and hourly data." Moreover, inserting a line break before this sentence would improve the readability of your conclusion section.
Thank you, we will follow your suggestion.

Line 507: "The strongest positive trends in daily 20-year return levels tend to occur in the shoulder seasons in the south". Please rephrase: "We find significant positive trends in daily 20-year return levels in…". This clarifies that the trends are actually high, and does not allow for speculation if the strongest trends of overall low trends are found there. Moreover, please explain what shoulder seasons are, or, since you do not use this term anywhere else, just name the seasons you mean explicitly. Lastly, from here on until the end of this section, it is not entirely clear what you mean with "summer" or any other season, if it is the traditional seasons or the ones you newly defined. Removing the traditional seasons from the figures would help clear this up as well.
Thank you for your suggestions. As discussed earlier, we will focus only on the new seasons in the revised version of the manuscript, thus removing any ambiguities in the text.

Line 509 ff.: You do not explicitly mention the June hourly trends anywhere else in the manuscript, it comes as a bit of a surprise here in the conclusion. Consider mentioning this already in the results and discussion section.
We will include this in the results (l. 365) and discussion sections (l. 459ff).

Line 509 (continued): You say: "Trends of hourly extremes are overall small, with the most notable exception being the month of June where we find locally high positive trends. In summer, hourly trends are almost systematically higher than the daily trends,... " Please clarify if trends in hourly extremes are now overall small or systemically higher than daily trends. You could also replace "summer" with "June", to avoid this conflict. Moreover, please explain how

you identified that hourly trends are “higher” than daily trends, because a comparison of absolute values, as you show in your figures, does not make sense, and you do not show percentage changes in the manuscript.
We acknowledge that our formulation is unclear, and we will rephrase it in the revised version. We mean to highlight here that daily trends are often negative, while hourly trends are around or above zero, thus ‘higher’.

Line 511: “...which leads us to state that a decoupling is taking place between processes generating extremes on the different timescales. While a general drying of the region takes place in summer, higher potential energy as a result of a temperature increase may lead to an increase in occurrence and/or intensity of short-duration convective precipitation events.” Were the processes causing hourly and daily extremes ever coupled? What you describe is also not a decoupling but simply different drivers developing differently. Suggestion to simplify this: “An explanation of why summer daily extremes seem to become weaker while hourly ones become stronger, could be that a general drying of the region takes place in summer, limiting the overall potential for intensive precipitation over a longer time scale. At the same time, higher potential energy as a result of a temperature increase may lead to an increase in occurrence and/or intensity of short-duration convective precipitation events.”
Thank you for your suggestion, it indeed improves clarity.

Line 513 f.: "This can especially be seen in early summer (June), when there is still enough moisture available for local recycling in convective processes." This can not be seen in your data as you do not show available moisture. Please provide a reference for this or clarify / better support this argument.
Thank you, we will expand on this.

Line 516 ff.: “Overall, the increases we find are strongest for daily extremes in autumn (October to December) and in June and November for hourly extremes. On the other side, September is the month experiencing strongest negative trends of extremes (although less for hourly extremes), which we attribute to an increase of aridity in summer and a later onset of the rainy autumn season as compared to the past.“ Please consider inserting another line break before this part. Moreover, similar to my comment to line 507, consider rephrasing this to be less ambiguous. Lastly, I recommend giving a more high-level summary here instead of reiterating on these specifics. What are the main take away messages of this work? For example, that considering a shift between seasons is essential to identify trends in extreme precipitation? Or that you do not find consistent trends over the whole region because underlying drivers of extreme precipitation differ between time scales, seasons and the North and South of the domain? You could also highlight the “tug of war” between increased potential for convection and moisture limitation during summer.
Thank you for these suggestions, we will reformulate the conclusion and elaborate more on the implications of our study.

Technical corrections

Line 6: in situ → in-situ
Line 13: driven by → caused by
Line 25: In October 2024?
Line 55: Given that in smaller catchments,
Line 152: "As in the wider region of the north-western Mediterranean" → "Similar to the wider region...."
Line 173: "registered by mechanic (past) or digital (tipping gauge, today) means" → registered by mechanic (past) or digital (today) tipping gauge.
Line 207: "for a given season" → for a given season/month
Line 230: "to characterize the general patterns" → please be more concise
Line 295: "sometimes" → too colloquial
Line 296: "after some initial testing" → too colloquial
Line 296: "Also, ..." → too colloquial
Line 319 and 320: In the north, and In the south, to avoid confusion (could read “the North Atlantic…”
Line 335: "stations is" → stations are
Line 431: "the ~~the~~"
Line 435: "the higher ~~amount of~~ humidity"
Line 437: "(6)" forgot "Fig." ?
Line 459-461: Please split up this long sentence for better readability.
Line 471: "We ~~could then~~ further hypothesize that"
Line 498: "Regarding the spatial and temporal ? of extremes...." forgot a word?
Line 517: “On the other side,” → on the other hand,

Thank you for these corrections, we will include them in our manuscript.