Review of “Climate variability can outweigh the influence of climate mean changes for extreme precipitation under global warming”

General comments:
This paper investigates the change in number of extreme precipitation and daily maximum temperature days at various global warming levels and time periods in response to various forcings. The authors show the increase in extreme precipitation days is generally associated with changes in the shape of the distribution whereas daily maximum temperature is generally associated with a shift in the distribution. The details of the spatial pattern of these responses depends on models, possibly due to differences in representation of aerosols and climate sensitivity.

The analysis and findings presented are interesting and a useful addition to the literature on the response of precipitation and temperature extremes to various forcings. A comprehensive analysis like this is useful and I think the manuscript can be further improved if the authors better motivate the reasons for looking at different global warming levels, SSPs at different decades, and different forcings. Specifically, while the current introduction describes the literature in detail, it lacks a clear statement on what is unknown and a research question the manuscript addresses. I suggest the authors add a paragraph in the introduction dedicated to this. I also have several specific comments below that I would like to see addressed prior to publication.

Specific comments:
Line 65: What is the reason for choosing JJA instead of looking at JJA in the NH and DJF in the SH? The subsequent figures show that several responses are hemispheric asymmetric (e.g., wetting in Northern Africa vs drying in Southern Africa, role of mean change in extreme precip appears mostly in the Southern Hemisphere). Is this asymmetry due to the seasonal asymmetry in NH and SH? I think a discussion on this and showing the DJF equivalents of the main figures in the appendix would be important.

Line 81-83: I think it would be better to introduce the definition, “PDF of the total change”, after the discussion of removing the annual cycle. I expected the leftmost panel of Fig. 1c to be a PDF of temperature before removing the annual cycle because the sentence describing the removal came after the definition of the “PDF of the total change”.

Line 81-86: This section is confusing to read because of the lack of detail of the kind of mean or variability the authors have in mind. For example, in the sentence “The second step involves removing the annual cycle at each grid point for each GWL which gives a PDF that only differs in variability”, I believe the authors are saying that by removing the annual cycle, the resulting PDF quantifies *daily* variability. It would be helpful to the reader if the frequency or spatial details of the mean and variability are specified here and elsewhere in the paper. If the same kind of mean or variability is used hereafter, the authors can define this once here.

Line 86: How are the changes in standard deviation and skewness isolated and quantified?

Fig. 1: Considering that the paper focuses mostly on precipitation, would it not be preferred to show this example using precipitation?

Fig. 1b: Is this for an example grid point or the global mean? What is the location? What do the different colored lines represent? Do they represent hot, median, and cold percentile temperature? Can you add a legend or describe it in the caption?

Line 126: Is this because there is only one ensemble member per model for PDRMIP simulations? If so, it would be helpful to state this explicitly here.

Fig. 2: Is there a reason for highlighting these 3 models in particular? NorESM1 appears to be highlighted because the response to CO2 is the smallest but the reason for highlighting the other two models is not provided.

Line 144: What is the significance level and is the statement here that the spatial patterns of changes in SD and number of extreme days significantly different or correlated?

Line 146-147: I’m confused by this interpretation. How does a higher correlation between changes in SD and extremes evaluated separately for different forcings tell us how the PDF change is more dependent on changes in aerosol vs CO2? Since SD is influenced by changes in both low and high precip extremes, my interpretation of this result is that the response to

aerosol forcing disproportionately affects high vs low precip extremes and thus is better captured by SD, whereas the response to CO2 forcing may also be affecting low precip extremes.

Fig. 3 caption: Do you mean $p < 0.05$?

Line 165: It looks like an increase in the number of intense precipitation events due to daily variability change is not seen everywhere in Asia. For example in MPI-ESM1-2-LR, there is a decrease along the 30th parallel north and for ACCESS-ESM1-5 and CanESM5, there is a weak response in the Middle East and Central Asia. I suggest you specify the subregions within Asia where the response is large.

Fig. 4 and 5: Is there stippling anywhere in these plots? I don't see them even when I zoom into individual subpanels.

Fig. 5: It would be easier to compare the role of mean vs variability change if the same colorbar settings were used for both Fig. 4 and 5. Currently the colorbar in Fig. 4 saturates on the positive end at 5 whereas the colorbar in Fig. 5 saturates at 1.

Line 175 and Fig. 6b: I'm surprised the change in mean dominates in Eastern Brazil, Southern Africa, and Northern Australia. Looking at Fig. 4 and 5 it appears to me that the brown is a darker shade in Fig. 4 compared to Fig. 5 for both MPI and CanESM. Can you provide plots with the variability and mean components side-by-side?

Line 181: What does "rarely observed" mean? Is it the 0.999 quantile?

Line 188-191: So the results shown here are not following the methods described in line 87-88? If so, you should mention in the methods section that you consider two different quantiles of extremes for CMIP6 data.

Line 201-202: Since not all of the domain shows a decrease in days of extreme precipitation I suggest you specify the region (e.g., South, Southeast, and East Asia)

Line 203-204: What you are describing here is only apparent around the Tibetan Plateau. I suggest specifying this.

Line 205: What does significant mean here? Statistically significant? Significant for impacts?

Line 222-234: Do you follow a systematic process for labeling a distribution as Gaussian, Gamma-like, or exponential-like? Otherwise characterizing the distributions like this is subjective and arbitrary. What are the standard deviation, skewness, and kurtosis of these distributions?

Line 231: Have you quantitatively tested the similarity between different distributions? (e.g., Kolmogorov Smirnov test)

Line 235-239: Can you discuss the regional dependence of the relative role in SD vs skewness change? For example in MPI, SD change is particularly large in WAF. In CanESM, SD change is large in SAS.

Technical comments:
Line 27: "unceratinties" → "uncertainties"

Line 121: "asses" → "assess"

Line 144: "test test" → "test"

Line 160 and 175: "South Africa" → "Southern Africa"; South Africa is a country whereas Southern Africa refers to the southern half of the African continent. I assume you mean the latter vs the former.

Line 171: "northern America" → "North America"

Fig. 9 colorbar label: "extream" → "extreme"

Line 257: "underplaying" → do you mean "underlying"?

Line 282: remove "net"