R#1 answer to comments:

Firstly, we would like to thank the thorough comments from Referee #1 as we believe they have added a lot of value to the manuscript. The following is a point-by-point answer to its comments:

The authors have achieved a thorough revision of their manuscript which is now in my opinion suitable for publication after minor revisions. Below are a few minor details.

The only main comment I have concerns the definition of significance for the multi-model mean change (related to one of my initial comments). I remain unconvinced by the statistical soundness of the authors' approach. The t-test for the difference in mean between two samples requires that each sample shares the same distribution, which for the future scenarios is clearly not the case. Inter-annual variability and temperature sensitivity to GHG forcing are different from model to model. Consequently, it doesn't make sense to define a significance level for the multi-model change as is done here. Instead, it should be evaluated on a model-by-model basis (and pooling the results together by — for instance — requiring a minimum number of models exhibiting significance). Don't get me wrong, it still makes sense to compute the multi-model average, of course, but that is not the same as attributing a degree of statistical significance to its change. This remains however a minor point of the paper and it does not modify the conclusions, so I would be fine with leaving it as is, though I think the authors should add a small qualification around ll. 157-158 to clarify the points I raised above.

Thanks for clarifying the issue you see in the test as we have applied it. To be more consistent in our significance definition, we have clarified the assumptions made to use the t-test as applied in our work. The assumption is that the variances from both reference and future period distributions are equa. Nevertheless, as you mentioned, it should not change the conclusions of the manuscript.

Changed:
A change in the multi-model mean is considered significant when it is beyond the threshold of a two-tailed t-Student test at the 95 % confidence level. The historical and future ensemble mean change and their inter-model standard deviations are used to compute the t-statistic.
to:
A change in the multi-model mean is considered significant when it is beyond the threshold of a two-tailed t-Student test at the 95 % confidence level. We consider that the null hypothesis is met when there is no difference between the multi-model distribution in the reference and future periods, presuming that the variability is stationary and present and future distributions are similar. To compute the t-statistic, first, each model's mean is computed from its members, and secondly, the multi-model ensemble mean and standard deviation are calculated.

Small comments

l. 6 maybe "in the multi-model ensembles" instead of "of the multi-model ensembles"

Suggestion applied.

ll. 14-15 "While there is less disagreement in projected precipitation between CMIP5 and CMIP6" -> you mean less disagreement than for temperature, correct? Maybe consider reformulating to make it explicit.

It has been specified that we are talking about less disagreement with respect to temperature.

ll. 124-125 "to calculate each of the model's and observational dataset's trends" -> "to calculate trends in each model and observational dataset"

Suggestion applied.

l. 125 "dependant" -> "dependent"

Suggestion applied.

l. 136 "the whole ensemble differences" -> maybe "differences for each ensemble member"

Thanks to this comment an error in the description of the Mediterranean hotspot has been found and solved (l.13*,l.136).

ll. 204-206 "The spread of the multi-model ensemble trends contain the observational ensemble trends (see Fig. 1). Mostly, for seasons SON and MAM, the observations fall inside the 90 % spread of the multi-model ensemble historical runs (not shown)." -> this is not extremely clear, because in the first sentence you do not specify that you are talking about DJF and JJA only. I also don't understand why SON and MAM are different; for DJF and JJA (Fig. 1) observed trends also fall within the multi-model 90% range. You could reformulate in a single sentence, like "PR and TAS trends in the observational ensemble fall within the range of the multi-model ensemble in all seasons (see Fig. 1 for DJF and JJA results)".

The suggestion has been applied, additionally specifying that results from MAM and SON aren't shown.

l. 217 "winter, summer" -> it might be better to stick to season acronyms throughout the paper (DJF, JJA, etc.) Right now, you keep jumping from summer/winter to JJA/DJF and vice-versa. (just a suggestion though)

The authors agree it's better to stick to either summer/winter or JJA/DJF. We have switched to acronyms after defining them in the methods section.

l. 253 "The results from this figure" -> I would repeat "Fig. S4" for clarity.

Suggestion applied.

l. 256 "the low emission scenario panels" -> referenced figure?

Labels have been added to the figure and are now referenced in the text after mentioning the low emission scenario panels.

l. 271 "find" -> "found"

Suggestion applied.

l. 301 "The rest of seasons" -> "MAM and SON"?

Suggestion applied.

l. 302 "Nevertheless, during the 21st century under the low emission scenario a slight increase in mean winter precipitation is projected" -> in MAM and SON? Please specify.

Suggestion applied.

l. 324 "the observed winter precipitation variability in the time series" -> "the inter-annual variability in observed precipitation time series" Also, this statement is a bit confusing: how do you see that inter-annual variability in models is lower? Did you check for each individual model?

This is an unsupported inference we made from the inter-model spread. Therefore, the authors have decided to erase the last part of the sentence.

l. 383 You might also take a look at Boberg and Christensen (2012) for the CMIP5 case https://doi.org/10.1029/2012GL053650

Thanks for sharing this work with us. While it is a very interesting publication, the authors consider it is out of the scope of the discussion about how the regional-global projections in CMIP6 are not amplified compared to CMIP5. Even if we haven't implemented the methodology to correct the temperature-dependent biases, we recognise it can be a relevant in-depth study to be considered in the future.

l. 384 "stuided" -> "studied by"

Suggestion applied.

l. 387 Brogli et al. (2018) argued that "a poleward expansion of the Hadley cell is of minor importance for the Mediterranean [temperature] amplification".

You are right. It was a typo as the text "(summer PR, winter and summer TAS)" should go after the lapse-rate argument, and the Hadley cell shouldn't be mentioned. The text has been corrected.

l. 443-445 This statement is a repetition from the end of part 3. You could leave it out here.

Suggestion applied.

l. 452 "The amplified warming of the Mediterranean especially affects temperature during summer and not in winter" -> "The amplified warming of the Mediterranean is found in summer and not in winter."

Suggestion applied.

l. 454 "no enhanced warming" -> specify "no enhanced regional warming" since at the global scale, CMIP6 is indeed warmer

The authors agree on the suggestion.

R#2 answer to comments:

The authors would like to thank Reviewer #2 for all the suggestions made throughout this revision, which helped improve the quality of our work.

The authors have done a very good job revising the manuscript. Yet, some minor revisions are needed:
Lines 43-45: You still have to state that the different ensemble members of a given model result from perturbations to initial conditions? Perturbations to model parameterizations? Running the model several times for the same scenario will lead to an ensemble of realizations only if some changes exist in the different members. Please, look at the experiments design of CMIP5 and CMIP6 and shortly write what perturbations were done.

The members differ in their initial co at the beginning of the historical run, either by starting from different years in the control run or by introducing perturbations in the last control run timestep. We haven't entered in such detail in the revised text but we have added that members are obtained from differences in the initial conditions (l.43)