Answers to Referee#3

General response:

The authors would like to thank the referee for the time and effort put into this review. These comments have been useful in improving our manuscript. We have carefully read the comments and provide a detailed response to each comment.

This paper presents results from a regional dynamical downscaling of two GCM ensemble from the HAPPI project and compares the model output to 4 extreme event indices.

The large number of ensemble members and important can potentially make a good contribution in how to analyse and estimate the difference between relatively close climate change "targets".

One major discrepancy however is that there are no evaluation of the quality of results with respect to observations or reanalysis. If the model results deviates to much from the "real world" one can not trust the conclusions given for the warm periods. The discussion does not have to be very advanced but I think you should include a point on this.

Answer: We did a quick analysis on this and found that the results are in good agreement with other downscaling activities using REMO for Europe (e.g., CMIP5 downscaling of GCMs). As suggested to referee#2, we will add a short analysis as supplement.

In general both the description of methods and results are not always very clearly defined. Sometimes it is just a matter of language. The sentences are sometimes so long that the reader loose the thread before getting to the end, so please try to be more focused.

A: We re-formulated many sentences that caused confusion in response to Reviewer#1 and #2.

A problem is that it is sometimes unclear whether the assumptions used are your / RCM limitations or inherited from the Happi protocol. Event hough I know the Happi protocol it was sometimes hard to separate. I see that many of the questions from other reviewers on critical assumptions e.g. how does pre-industrial come into play, when is the future time slice .. is often related to the protocol. I think the authors should include a summary of this information, in particular as you rightly point out that the Happi protocol is quite different from the traditional CMIP settings.

A: We came to the same conclusion. Therefore, we separated the description of the HAPPI protocol from the REMO model set-up and extended it (see answer to Reviewer#1).

The defintions/ presentations of the indices and how they are used can be improved (Section 2.2 and 2.3) It is particularly hard to follow the setup of significance tests. "RX5 days are computed ...similar to ATG28, however …

A: We will adjust the text and add sub-sections for each indicator.

CDD is similar to ATG28 but with the method used for RX5 but now however? There are no significance test for RI50yr ? There does not have to be one but I would like to know.

A: No, we did not use a significance test for RI50yr.

With only 4 indices I think you should avoid "similar" as much as possible and just describe the methods for each.

A: Separating the description of the indicators should make it more clear (see answer above).

The discussion on being able to reproduce the more noisy results of the smaller ensemble by picking a smaller number of the largest ensemble was interesting. I wonder if it is any way of presenting this in a cumulative manner, with respect to number of ensemble members. The change rate may also indicate whether even the largest ensemble is too small. I realize however that this is likely beyond the scope of the article.

A: We will include an analysis on the randomly picked 25 member ensemble from ECHAM6 as a supplement as suggested to Reviewer#2. We agree the further analysis on the change rate would be beyond the scope of this article.

To the point on data availability: Is it the data used specifically for this study, or is it general RCM output.

A: We are still struggling with finding a data centre that can host the data. Due to CMIP6 activities our dataset is of low priority. However, we managed to cmorize the most common variables and can provide them by request. This is already ongoing in the framework of other projects we are working on.

Some minor suggestions /corrections

Line 20: "differ" not needed

A: We will rephrase this sentence to (see also answer to Reviewer#1):

"Identifying regional climate change impacts for different global mean temperature targets is increasingly relevant to both the private and public sectors."

Line 25: Include a line on scenario definition in CMIP6?

A: We will add CMIP6 to this sentence:

"Temperature targets, however, are not directly related to the Representative Concentration Pathways (RCP, Van Vuuren et al., 2011) used in the experimental design of CMIP5 (Taylor et al., 2012) or the Shared Socioeconomic Pathways (SSP, Meinshausen et al., 2019) used in CMIP6 (Eyring et al., 2016)."

Eyring, V., Bony, S., Meehl, G. A., Senior, C. A., Stevens, B., Stouffer, R. J., and Taylor, K. E.: Overview of the Coupled Model Intercomparison Project Phase 6 (CMIP6) experimental design and organization, Geosci. Model Dev., 9, 1937–1958, https://doi.org/10.5194/gmd-9-1937-2016, 2016.

Meinshausen, M., Nicholls, Z., Lewis, J., Gidden, M. J., Vogel, E., Freund, M., Beyerle, U., Gessner, C., Nauels, A., Bauer, N., Canadell, J. G., Daniel, J. S., John, A., Krummel, P., Luderer, G., Meinshausen, N., Montzka, S. A., Rayner, P., Reimann, S., Smith, S. J., van den Berg, M., Velders, G. J. M., Vollmer, M., and Wang, H. J.: The SSP greenhouse gas concentrations and their extensions to 2500, Geosci. Model Dev. Discuss., https://doi.org/10.5194/gmd-2019-222, in review, 2019.

Line 33:skip the word "indeed"

A: We will change the sentence to:

"The high natural variability in models requires the creation of large ensemble datasets (Deser et al., 2013)"

line 70: Green-house-gas -→greenhouse gas

A: This will be changed accordingly.

line 81: Sea-Ice → sea-ice (usually not capital letters)

A: This will be changed accordingly.

line 121: Missing intervall ("100" years on both sides)

A: This is a typo and should read "between 10 and 100 years".

Figure 1. Is the outer map equal to the model domain?

A: Yes, the outer domain is equal to the model domain.

Line 135-136 The explanation is fine, but then you do not need to mask it out either. It would show up as insignificant?

A: We agree that masking "non-significance" and masking for other reasons is confusing. We will change that in an updated version of the Figure. We masked the SST before running the test. Hence, we do not know if it would show significance or not.

Line 141 NOResm → NorESM

A: Will be changed accordingly.

Figure 6. The difference between the two downscaling sets are quite large. Any comments.

A: This is related to the forcing model. On the one hand, the NorESM ensemble is wetter compared to the ECHAM6 ensemble. On the other hand, the ECHAM6 driven ensemble seems to be biased towards warm/dry conditions. This can be examined in more detail when a small evaluation is added to the supplement.

line 260 -265 use significance instead of unlike / similar

A: We will change the text to the following formulation (see also response to referee#2):

"The changes to CDD distributions show that Spain will experience significantly more drought conditions in the future compared to the current period, even at a 1.5°C increase in GMT. For Italy, drought conditions associated with the 1.5°C simulations show non-significant changes, yet those associated with the 2.0°C simulations are significantly different to the current period, thus showing possible consequences of exceeding the 1.5°C GMT target of the Paris agreement."