RC2#1: The submitted paper holds a thorough analysis of seasonal mean changes in European temperature and precipitation for the mid-century and end-of-century. It utilises the CMIP5-based EUROCORDEX multi-model GCM-RCM experiments, and a specific method to estimate the so-called 'balanced' climate change response and 'balanced' uncertainty. This uncertainty is then attributed to GCM, RCM, RCP (or interactions between those, or internal variability.

This is an important exercise, and hence I am of the opinion that the paper contributes usefully to the existing scientific literature. I recommend acceptance of this manuscript for publication, though would recommend major revisions to the text before doing so. I will try to explain my discomfort with the text in its current form below.

We thank the reviewer for this positive feedback and for these numerous constructive comments that will certainly help to improve the manuscript.

RC2#2: As noted, the results of the study are important, and go beyond existing work and published analyses of EUROCORDEX as far as I know. However, given the conclusions and the recommendations made there, I would like to see much more comparisons between the QUALYPSO method and the 'normal' approach. This to inform the reader of the gains made and the errors that could otherwise be introduced. I also wonder if a sensitivity analysis, e.g. by splitting the ensemble in two smaller ensembles, computing two times the balanced response and the normal response, and showing (rather than telling and trusting) that this is indeed a robust method.

We fully understand this concern and this type of comparisons was actually considered in a first version of this paper. An additional comparison has thus been carried out and will be included in this paper. An important aspect of this sensitivity analysis should be to keep the same set of RCMs/GCMs since, as shown in the paper, the overall contribution of GCM/RCM uncertainty is greatly dependent of a few RCMs/GCMs, regardless of the approach for partitioning the uncertainties.

To perform this comparison, we rely on a complete synthetic MME composed of 9 GCMs x 13 RCMs x 3 RCPs = 351 climate change responses generated using ANOVA effects and residual variability estimated with the original MME. We then subsample randomly 1,000 different MMEs of 87 chains among this complete synthetic MME of 351 chains. Figures 3 and 4 below show the mean change estimates (BM and M) obtained for temperature and precipitation, respectively, for the different RCP scenarios, SREX regions and seasons. Clearly, the variability between the estimates M obtained with direct averages of chains available for each scenario is larger than when QUALYPSO is applied (BM estimates). For temperature, mean estimates obtained with QUALYPSO are particularly stable.

Note than many other experiments could have done. Here, we ignore internal variability, in order to minimize the computational burden (estimation of the climate change response). This random subsampling is obviously the simplest way to perform this experiment. More realistic subsampling strategies reflecting the overrepresentation of the scenario RCP8.5 or of some RCMs/GCMs have been tested and do change the conclusions. Similarly, tests with smaller subsampled MMEs lead obviously to larger variabilities between the different samples, which is relevant for other CORDEX domains often counting less simulations.

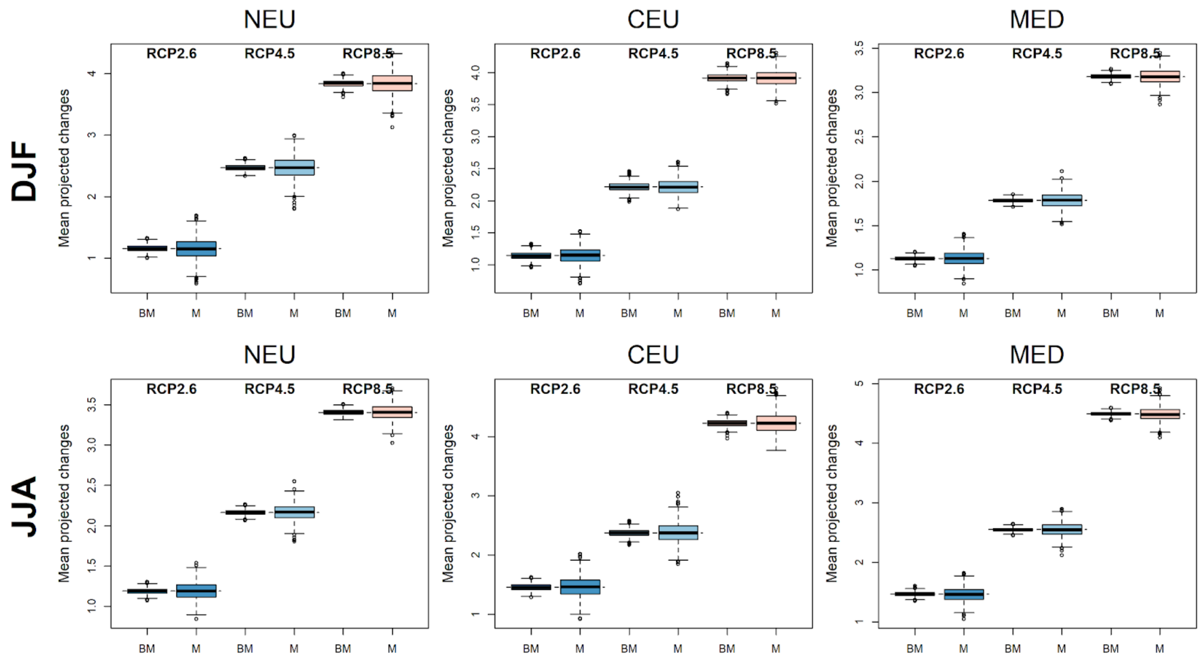

Figure 3: Comparison of mean projected changes estimates for temperature using QUALYPSO (BM) and direct averages (M) of a synthetic MME for each RCP scenario, SREX region, and season. A complete synthetic MME composed of 9 GCMs x 13 RCMs x 3 RCPs = 351 chains is generated using ANOVA effects and residual variability estimated with the original MME. The boxplots show mean change estimates based on 1,000 random subsampling of 87 chains among the complete synthetic MME of 351 chains. Dashed horizontal lines indicate the corresponding averages obtained from the complete MME.

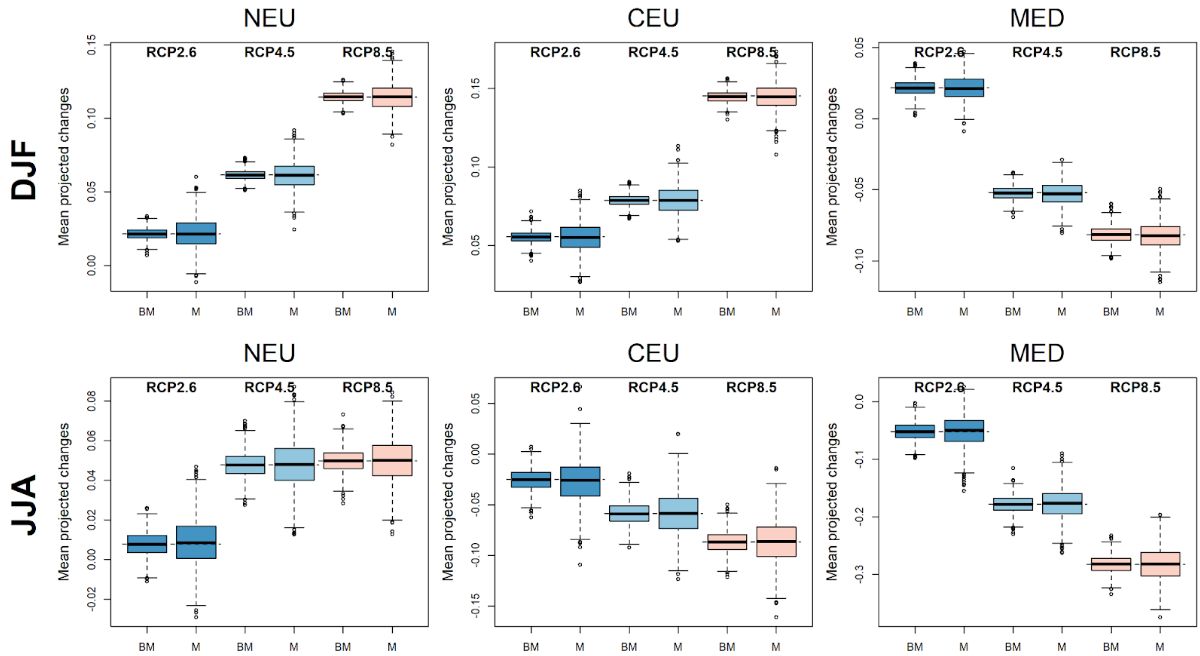

Figure 4: Comparison of mean projected changes estimates for precipitation using QUALYPSO (BM) and direct averages (M) of a synthetic MME for each RCP scenario, SREX region, and season. A complete synthetic MME composed of 9 GCMs x 13 RCMs x 3 RCPs = 351 chains is generated using ANOVA effects and residual variability

estimated with the original MME. The boxplots show mean change estimates based on 1,000 random subsampling of 87 chains among the complete synthetic MME of 351 chains. Dashed horizontal lines indicate the corresponding averages obtained from the complete MME.

RC2#3: In its present form, the paper does not excite the reader, or invite thorough reading. The method section is confusing, and does not offer a clear explanation of QUALYPSO to the reader:. I suggest a more understandable and intuitive explanation is added, maybe with some drawn schematics that show how differences between model runs would be quantified in the balanced outcomes and uncertainty quantifications. The results sections feel like a dump of many figures and tables, with lots of text that tell us what can be seen in the figures. It would be more useful to explain sources of differences (e.g. climate sensitivity of GCMs is only noted once in line 332!).

We appreciate this comment and will try to give a less technical presentation of QUALYPSO in the revised version of the manuscript. A figure illustrating the different steps will be added, similarly to the attempt made in Hingray et al. (2020). We will try, as much as possible, to simplify and shorten the description of the results in the revised manuscript in order to improve the readability of the paper.

Concerning the interpretation of the results, we agree with the reviewer that tracking the sources of the model differences is an interesting and relevant topic. However it likely requires a dedicated study for each model family, each variable, geographical region or season which is out of the scope of the current study. We therefore decided to rely on the published literature known by the authors to give some hypotheses. We acknowledge that this approach remains limited but we hope that further studies, starting from our assessment, may explore those differences in depth. In particular defining the equivalent of the GCM ECS but for the RCMs would be very relevant but does not fit in the current study.

References

Hingray, B., Evin, G., Blanchet, J., Eckert, N., Morin, S., and Verfaillie, D.: Partitioning uncertainty components of an incomplete ensemble of climate projections using data augmentation, EGU General Assembly 2020, Online, 4–8 May 2020, EGU2020-21864, https://doi.org/10.5194/egusphere-egu2020-21864, 2020.