Response to the comments from Anonymous Referee #1 for the manuscript:
"Uncertainties in climate change projections covered by the ISIMIP and CORDEX model subsets from CMIP5" by Ito et al.

We would like to appreciate your careful review and constructive comments and suggestions for improving our manuscript. We almost agree with them. We have made modifications through our manuscript according to the responses. Please check our detailed responses below. The numbers of page and line are corresponding to the number in the original file (https://www.geosci-model-dev-discuss.net/gmd-2019-143/gmd-2019-143.pdf).

In this modification, we added a CMIP5 model of CSIRO-Mk3L-1-2, to the original 49 models for the historical run we analyzed. It is because there is no member of r1i1p1 by CSIRO-Mk3L-1-2, but there is r1i2p1 as well as CESM1-WACCM which was already used. The results did not change from the original manuscript by this modification. I apologize for the change.

In this revision, McSweeney and Jones (2016) have referred to as MJ2016 except for the first reference.

Thank you once again for your review.
We would be glad to respond to any further comments you may have.

--- Summary and General comments

The paper by Ito et al. investigates the uncertainty ranges in projections from the ISIMIP and CORDEX projects. Both of these projects selected a sub-sample from CMIP5 Global Climate Models (GCMs) to bias correct and then drive impact models (ISIMIP) or to downscale the GCM's (CORDEX). ISIMIP and CORDEX have different goals and also the number of models selected and the approach to sub-select the GCMs were different. The authors look into how well these two projects cover the uncertainty ranges provided by the original CMIP5 model set. They show that the ISIMIP and CORDEX uncertainty ranges are smaller than the original range but still larger than from a subset only selecting well performing models, even though the number of models selected in ISIMIP and CORDEX was smaller than the number of well performing models they were compared to. The authors also conclude that better subsets with smaller biases and/or higher scores would be possible than the current ISIMIP and CORDEX selections.

While it is interesting to see how different the uncertainty ranges of different model selections are, I am not necessarily sure if the comparison is fair, given that as far as I know neither ISIMIP nor CORDEX selected their GCMs based on these criteria. Among other points explained below, I am also missing a clear recommendation that would help the next rounds of ISIMIP and CORDEX to sub-select their GCMs.

We glad to hear your interests in our study. We have made our response to the comments about the unfair comparison between the subsets from ISIMIP and CORDEX and about the recommendations towards the next rounds, as the responses to specific comment #1 and #4 respectively. Please find below.

--- Specific comments

1. For ISIMIP the main constraint in choosing GCMs was data availability, and they needed many more variables than the ones the authors consider in this study. Hence, even if "better" subsets in terms of performance based on precipitation and temperature would be possible, that does not necessary mean these subsets would have been an option for the ISIMIP project. For CORDEX data availability was also a major constraint, so again, even if better subsets based on temperature and precipitation would have been possible, if the data to drive the RCMs was not available that would not have helped the CORDEX project. These aspects should at least be discussed in the manuscript.

We appreciate your accurate comments. Our explanation was not sufficient. The "better subset" is based only on the model bias and Taylor's skill score in our analysis. From an additional analysis in this revision, it is found that such a subset can be obtained under the condition without considering the data availability and with focusing on one variable of temperature or precipitation. We have described the following sentence to the section of discussion which made in this revision.

(P8 L23) "… a much better model subset, regarding to biases and skill scores, can be selected with making use of the advantage of the small number of models. However, such a selection can be conducted when there are no constraints of data availability which was the main constraint to select the current subsets in ISIMIP and CORDEX and when we use one variable of either temperature or precipitation."

As you noted, ISIMIP and CORDEX select their subset under the different constraints at the present. We have also added the followings in the section of discussion:

(P8 L23) "In this study, we assessed the current ISIMIP and CORDEX subsets to investigate whether the subset indicates small biases in the historical climatology and covers the uncertainty in the future projections widely using temperature and precipitation. Both variables are most frequently used in future projections and also weather forecasts. The evaluation for such a principal variable is important for the studies of ISIMIP and CORDEX. It should be noted, however, that ISIMIP needs the dataset with reasonable for multiple variables used in their impact assessment and with enable to discuss the uncertainty in the projections. CORDEX requires the dataset with based on a plausible mechanism of the climatology as the input data for RCMs. Thus, there is a possibility that a good subset which we presented based on the model performance for temperature and precipitation will be an option of their future subsets."

2. I was also missing the link from the performance in the historical projections to the projected uncertainty ranges. Do the sub-sampling based on lower bias/higher score cover larger, smaller or similar uncertainty ranges in the projections? The data is all there in the figures, but it is not discussed in the text.

We had mentioned the uncertainty range for the temperature change obtained from the subsets on P6, L24-26 and on the other hand, for the precipitation change on P6, L30. Especially for precipitation, there was less explanation. We have added the description below to P6, L29. With the addition, we have modified a whole of the paragraph more understandably.

"The subsets of $\Delta P(CMIP'_{lowB})$ and $\Delta P(CMIP'_{highS})$ cover 70% and 60% of the full range of uncertainty from $CMIP_{Full\_Future}$ as the average over 14 regions, respectively, with totally covering the full range in Australasia. The largest difference between the coverages from $\Delta P(CMIP'_{lowB})$ and $\Delta P(CMIP'_{highS})$ appears in East Asia. Therefore, we need to pay attention that, when the model performance is the condition to select subsets, the uncertainty changes depending on which evaluation index are used, like at least the bias or the skill score."

3. I also find it hard to believe that neither the ISIMIP nor the different CORDEX regions did any analysis similar to what the authors provide here? At least for ISIMIP McSweeney and Jones (2016) seem to already have done this in a very comprehensive way. What is this study adding on top of that?

McSweeney and Jones (2016) (hereafter MJ2016) have discussed the uncertainty in the projections but not mentioned the ability to represent the present-day climate and the projections itself which we have investigated. Also, as the update from MJ2016, we have analysed four GCMs used in the newer round of ISIMIP, instead of the GCMs analysed in MJ2016. On the other hand, as you pointed out, there are some CORDEX regions where their GCM subsets have been assessed but the assessments are limited.

Uniform assessment over the regions permits to discuss the difference of performance among the regions. In addition, Gutowski et al. (2016) have mentioned there is a possibility of the heterogeneity on climate information among the regions as one of the main problems in CORDEX. This study has indicated that the subsets can widely capture the uncertainty in both projections of temperature and precipitation in the regions with a large ensemble. Thus, it is found the heterogeneity exists in the current dataset when focusing on the uncertainty. Furthermore, from the added results in this revision, we suggest that nine model members are needed to solve the heterogeneity of the uncertainty.

From the assessment of the subsets selected in each program in the same method, we understand how different the climate information from a global consistent subset is from the original one by using the ISIMIP subset in the CORDEX framework, with assuming CORDEX CORE.

We have added the above contents to P4 L3.

“The ability for the ISIMIP subset was not mentioned by MJ2016 and thus we investigated that in region by region. We analysed four GCMs selected in ISIMIP2b (unless specified otherwise, hereafter refers to as ISIMIP) here. Thus, assessment of the projections was also updated from MJ2016. The GCMs used in CORDEX have been assessed by region in previous studies, but are limited (e.g., Haensler et al. 2013 for Africa; Bartók et al. 2017 for Europe; Karmalkar 2018 for North America). Even simple assessment conducted is needed for the present CORDEX. Furthermore, uniform assessment across regions permits to discuss the difference of characteristics among the regions and the possibility of heterogeneous scenario as mentioned above. By using the subsets from the two programs, we can explore the difference between the original subset in CORDEX and the subset selected with assuming CORDEX CORE, which is helpful information for the model selection in CORDEX CORE.”

4. On page 8, lines 19-22, the authors mention results what would happen if a larger number of models would have been used in the Central Asia region. This result, I imagine something similar to Figure 3 in McSweeney and Jones (2016) but for the CORDEX regions, would have been very interesting. I think it would allow to show how many models would need to be selected to cover a certain uncertainty range, which would help to make a recommendation for the next round of CORDEX. I would also be curious to see if these numbers differ between different regions.

We appreciate your constructive suggestion to gain more insight into our results. We added the results about the change of coverage depending on the number of models in each region to Section 3.3. We have referred the idea by McSweeney and Jones (2016). They have changed the number of models to explore how the coverage changes with the number of models when a subset covers the uncertainty in each grid most widely over the globe or regions. On the other hand, in this study, to consider making better use of the current subsets, we have changed the number of models from the current model members and explored how the coverage changes. The details are as what followings:

(P8 L19) “From Fig. 4, the subsets with nine models can capture the uncertainty of projections in both temperature and precipitation widely, implying that there is a heterogeneity on the dataset by a different number of models (Gutowski et al. 2016). We explored whether a similar tendency can be obtained in the other regions when the number of models changed. The same approach was performed by MJ2016. They focused on a subset covering the uncertainty in each grid most widely over the globe or regions and investigated how the coverage changes with the number of models. On the other hand, in this study, to consider making better use of the current subsets, we investigated how the coverage changes with changing the number of models from the current model members.

Figure 5 shows the change of coverage performance with the number of models changing in each region. When the number of models is larger than the current number, we added models randomly selected to the current members. By contrast, when the number of models is less, we removed models randomly selected from the current members. Here we focused on the median of the FRA values obtained from the possible 10,000 random samples, meaning the FRA value obtained with a possibility of 50% when selected subsets randomly. For the temperature change, the median exceeds 60% in all regions when changing the number of models from the current four ISIMIP members to seven members which are less than nine members (Fig. 5a). The median above 60% is also obtained in 13 regions (except for Antarctica) when changing the number from the current CORDEX members to nine members. For the precipitation change, the coverage in nine members is above 50% in 10 regions and in 12 regions by changing the number of models from the current members in ISIMIP and CORDEX, respectively (Fig. 5b). Even when using nine members, the median is less than 50% in Four regions of MENA, Africa, and South and East Asia for the change of number from the ISIMIP subset and in two regions of MENA and North America for that from the CORDEX subset.

The IQR for ΔT shifts to a high FRA smoothly with the number of models in all regions. By contrast, the IQR for ΔP sometimes gets large suddenly and/or shifts sharply, for instance, MENA and Africa. The discontinuous change is caused by a large variance of ΔP from each model member. That is to say, when there are model members indicating a large change ratio relative to the other members, the coverage largely differs depending on the inclusion of the member with the large ratio or not. The change amounts, ΔT are similar among the model members and the variance is small. Thus, the FRA increases with the number of models and the IQR also increases smoothly. To prevent selecting the subset with a large change of the coverage depending on a model with extremely large or small change amount, investigating the variance of the projections in each region is needed when the number of models is decided."

(P9 L18) "The current CORDEX subsets can capture both uncertainties for temperature and precipitation in the regions with a relatively large ensemble. However, it is found that changing the number of models from the current CORDEX members to nine members can capture more than half of the full uncertainty in both projections of temperature and precipitation in more than 85% of all regions, with a possibility of 50%. Furthermore, the same is also shown as for the ISIMIP subset, but for 70% of all regions. Focusing the uncertainty in the future projections, this result proposes that the current number of models need to be changed to discuss a similar uncertainty range among the regions."

--- Technical corrections

1. Figures: While I kind of like the illustration of the graphs on the map it takes up quite a lot of space while the graphs itself are rather small. I wonder if the graphs could be increased but would take up less space in a more classical arrangement?

We appreciate your suggestion. We can understand a lot of space, especially in Supplement 4 and 5. We deeply considered the modification but the graphs on the map is good from the point of seeing the property corresponding to the region at a glance. We have redrawn the figures with reducing the space as much as possible.

2. Supplement 1: I find this table not very informative, I would be more interested to know in which regions which models were used than in how many regions each model was used.

The table has been changed to a table presenting the models used in each CORDEX regions. Please check the modified manuscript.

3. Supplement 4 and 5: I think the Obs are missing in these Figures.

Differed with the precipitation, one observation dataset, CRU, is used as the temperature reference data as indicated on P4, L26-27. Thus, there is no plot for the observation.

We have revised our manuscript to address comments from Anonymous Reviewer #1.