Response to reviewers
gmd-2016-170
Lynch et al.

We thank all referees for their thoughtful comments, which have substantially strengthened this manuscript. Per the _GMD_ instructions, we respond below to all comments and questions.

Original reviewer comments in italics. Responses in blue. Revisions in maroon.

Reviewer #2:

1. *Most of the results shown in this study are 30-yr averages. But at the same time the authors highlight the importance of this study for impact assessments. Are the authors able to demonstrate skill of the pattern scaling methods at impact relevant temporal scales?*

Response:
We did not make clear in the introduction that this analysis was intended to justify our methodology for creating patterns used in pattern scaling. We do not assess the skill of the patterns at impact relevant scales. We only examine mean annual, DJF, and JJA patterns. We have not analyzed monthly patterns in this manuscript. A discussion of this topic was added to the conclusion section of this paper.

---

2. *The model data was regridded to a T85 resolution (approx. 1.4 spatial resolution), even though Table 1 reveals that several models have resolutions which are significantly coarser than the resolution which is regridded to. Regridding to a finer resolution is strongly discouraged, as models clearly have no information at this scale due to parameterizations. This can lead to large errors in the model fields which is particularly harmful for the purpose of impact assessments. I recommend to regrid to the model with the coarsest resolution (in this case, MIROC-ESM which has a spatial resolution of 3o). Consider using the CDO function called remapcon (Jones, 1999) for this task.*

Response:
We agree that regridding to a higher resolution can be problematic especially when the variable analyzed has high spatial heterogeneity. Taking into account this issue we have regridded to the model ensemble's lowest resolution, which is MIROC-ESM. All figures were redone after regridding. A brief discussion of this was added to the "Climate Models" section.

Revision:
All model output were regridded to lowest spatial resolution of the ensemble prior to calculating ensemble mean or median. This was done for averaging purposes, as each model had a different spatial resolution. Regridding to the lowest resolution of the multi-model ensemble was necessary as regridding to the highest resolution of the multi-model ensemble would lead to errors if used for purposes of impact assessments.

---

3. *Having multiple models per institute is not the only reason why the assumption of model independence is not valid. An important point is shared code or even whole model components. See, e.g., Abramowitz, G., & Bishop, C. H. (2015). Climate model dependence and the ensemble dependence transformation of CMIP projections. Journal of Climate, 28(6), 2332-2348. The way the authors deal with the issue of model interdependence is not appropriate. Simply selecting the "best" model per modeling center does not get rid of the internal dependence structure. It is for example known that the ACCESS models are based on the HadGEM2 atmosphere and both of those models are part of the selected CMIP5 subset (see Table 1). See, e.g., Knutti, R., Masson, D., & Gettelman, A. (2013). Climate model genealogy: Generation CMIP5 and how we got there. Geophysical Research Letters, 40(6), 1194-1199. Model selection based on performance only is not recommended as one should ensure a high degree of independence at the same time. Pairwise error correlation can for example be used as a metric for model dependence, as used in Bishop, C. H., & Abramowitz, G. (2013). Climate model dependence and the replicate Earth paradigm. Climate dynamics, 41(3-4), 885-900.*

Response:

This is a good point. For this analysis we did not want to use a large ensemble of models, and we also did not want to arbitrarily choose a multi-model ensemble that would over-represent particular sub-models (example from above: Met Office Unified Model atmospheric model). However, we did want to choose a 10-12 model ensemble that realistically produced mean observed spatial/temporal climate. The metrics used were very simplistic and served primarily to exclude models that were statistically (we used RMSE) different from the observed data. (We stress that accurate simulations of global mean temperature do not mean that a model is 'better' or that 'good' models will have better predictive skill.) It is true, however, that we cannot say that we have achieved true model independence; we have included a statement in the Supplementary Material to convey this.

---

4. *The authors decided to use NCEP1 for the validation of their model ensemble based on annual and seasonal climatology. For proper model validation, I highly encourage the use of more than one observational/reanalysis product as observation products can vary significantly, even for surface air temperature. Especially in this case, I would not recommend the use of NCEP1 (a 1st generation NCAR product) as more recent reanalysis products are available. Here are some examples: JRA55, MERRA2, CFSR (3rd generation NCAR product), ERA-Interim. Why did the authors choose a reanalysis product over a gridded observation product? There are several gridded global observations available, especially for monthly temporal resolution. For example: Had-CRUT4, Berkeley BEST, GISTEMP, NOAAGlobalTemp. Reanalyses should only be used for evaluation if no gridded observation products are available, because they contain model output. Therefore, the model ensemble might not be truly independent of the chosen reanalysis product and a comparison is not fair.*

Response:

Validation of models was not intended to be the focus of this paper, but merely as a means to 'cull' all available models into a smaller, manageable ensemble. We agree that there are many observational and reanalysis datasets to choose from. For this analysis we used NCEP1, which has a longer time span (1948-present vs. 1979-present) which is preferable when constructing bias, standard deviation, and root mean square error.

A drawback of using gridded observations is that they do not account for the contribution to a regional average of areas with very few weather stations, especially high terrain. Reanalysis models assimilate observed data into a weather prediction model at the spatial resolution of climate models, and they process observations in the same manner as climate models, and arguably offer the fairest comparison with climate models (Mote and Salathe, 2010).
The description of NCEP1, rationale for model culling, and Figure 1, have all been moved to the Supplementary Material section. We believe that this is a more appropriate section for this information.

---

5. *The conclusion section does not contain any comments on the limitations/potential caveats of this study and on future work. Please add a few lines.*

Response:
Yes, this is needed. The conclusion section has been edited to include broad limitations and future work. In the conclusion we discuss impact relevant timescales, precipitation pattern scaling, and scalers from our SCM, Hector. It should also be mentioned that in the Results section, we have added a discussion subsection under each topic where we discuss caveats and broader issues.

---

6. *Abstract The abstract assumes that the reader already knows what the linear regression and delta methods are. Please add a 1-2 sentence explanation of each method as it is otherwise hard for the reader to understand what was done.*

Response:
Yes, this was not clear. The abstract will be edited.

Revision:
The delta method is simply the ratio of local climate change to global mean temperature change over two epochs, and the regression method uses global mean temperature as a predictor of local climate.

---

7. *The last sentence of the abstract says that the patterns will be used to examine feedbacks. Where in the main text are feedbacks discussed?*

Response:
We may have been too far reaching, but we do intend to examine feedbacks and uncertainty once we scale the patterns with a SCM in the future. The aforementioned sentence will be edited.

---

8. *Introduction Page1, Line 22: Typo: "It is used to..." rather than "It can is used to...".*

Response:
Yes, thank you. We have edited the introduction paragraphs and this line has been edited.

---

9. *Page 2, Line 3: No need to introduce the abbreviation IAM as it is not used later on in the article.*

Response:
Yes, thank you. We have edited the introduction paragraphs and this line has been removed.

---

10. *Page 2, Lines 8-10: The word "increase" is used 3 times in this sentence. Try to change that (e.g., use warming rather than increase).*

Response:
Yes, thank you. We have edited the introduction paragraphs and this line has been edited.

---

11. *Page 2, Lines 10-11: Please clarify this sentence. The pattern obtained with the pattern scaling methods is just a mean pattern and contains no information about the variability. This sentence therefore does not make any sense to me.*

Response:
Yes, thank you. We have edited the introduction paragraphs and this line has been removed.

---

12. *Page 2, Line 26: Please elaborate on why the reference period is expected to matter for assessing future change. Isn't it rather the length of the reference period that is important?*

Response:
Which epoch is used and the length of the epoch are both important. When using a pre-industrial reference period, the length of the epoch is not likely to matter. For the historical period, the climate models do show more warming in the late 20th Century as compared to the early 20th Century. However, we do not explore how epoch length may or may not change the pattern. This issue is not suited for the introduction, and a small discussion of this topic is included in the Methods section to justify our use of the standard 30-yr epoch.

---

13. *Data/Methods Page 4, Line 1: Which and how many ensemble members per climate model were used for the analysis?*

Response:
Thank you. This was not mentioned, and it should be noted as some models had multiple realizations. A sentence will be added to that paragraph (line and page numbers no longer match up to original manuscript).

Revision:
We used the first realization from each model, choosing not to average over multiple realizations from each model.

---

14. *Table 1: Add information about the chosen ensemble members. Replace "HadGEM-ES" with "HadGEM2-ES".*

Response:
Yes, thank you. This has been fixed, and per last comment, a sentence regarding ensemble member has been added.

---

15. *Page 4, Lines 8-10: Elaborate why the ensemble mean is not expected to be skewed when using model output that is not independent. Model replication automatically skews the mean of a distribution towards the most replicated one.*

Response:
We have moved this section to the Supplementary material and removed this sentence, but we would like to clarify it. For the analysis we culled the models to limit model bias, and we created an ensemble that attempts to limit model dependence (i.e. shared code is limited). So, in this case, with a purposely selected ensemble of models, the model bias is not thought to skew the resulting mean patterns.

---

16. *Page 4, Line 11: Many impact studies need precipitation information as well as temperature information. Could this analysis be expanded to include total precipitation rather than a single variable?*

Response:
For this study, we are only looking at temperature patterns. We fully intend to examine precipitation patterns and impacts after scaling these patterns with output from a SCM, Hector.
A sentence about future work has been added to the conclusion section.

---

17. *Page 4, Lines 14-15: The authors stated above that the assumption of model independence within the CMIP5 archive is not valid. In this case, how do the authors justify the use of the model spread as a measure of uncertainty? Having dependent models automatically leads to a bigger model agreement, which should not be regarded as an indication of robustness.*

Response:
We do recognize that an assessment of model uncertainty using a limited number of models will not be very robust. Because we are using an ensemble, some measure of uncertainty quantification must be done, and we use the Student's t-test as such a measure. That being said, the model selection attempts to achieve a small measure of model independence.

Revision:
In the Results/Discussion section:

We recognize that the use of the Student's t-test probability distribution has very limited means of evaluating significant differences between ensembles from each pattern methodology. As such, our treatment of model uncertainty is not robust. Ideally, a much larger ensemble would be used, which would allow a more rigorous testing of differences between patterns from each method and scenario to be done.

---

18. *Equation 1: It would be helpful for the reader to add the dimensionality of each element of the equation. So, DP and TL are 2D fields whereas TG is a number. The same could be done for Equation 2.*

Response:
Yes, thank you. Dimensions of input/output data in Methods section has been added.

---

19. *The authors calculate the delta pattern for each model and then use the ensemble mean /median for their analysis. It would be interesting to compare those results from the pattern obtained when using the TL and TG of the mean/median of the ensemble (instead of using the individual models and then calculating the mean/median).*

Response:
We did this calculation and have the plots for rcp8.5 and rcp4.5. For the rcp8.5 scenario the differences in ensemble averaging post and prior to pattern creation are minor for the delta and LSR patterns (less than 0.02 degC/degC). For the rcp4.5, the delta pattern difference is less than 0.035 degC/degC), but for the LSR pattern the difference is no larger than 0.15, mainly in the midlatitude Northern Pacific Basin, which is interesting.
We have a figure examining this difference, but it was tangential to the main points, and as such not included in the text.

---

20. *Page 5, Line 9: What is a gridded time series? Is that simply 3-dimensional data (time x lat x lon)?*

Response:
Yes. We did not make that clear. The revision recommended in Question#18, addresses this issue.

---

21. *Equation 3: Why is the "^" symbol on top of being used? Should this also be done in the text which describes this equation? Does "n" refer to the number of time steps? Please add that to the description of the equation.*

Response:
The circumflex is often used to denote unit change, but it was not discussed in such a way in the text. The 'n' is indeed number of time steps. A line was added below the beta equation to define both of these issues.

---

22. *Page 5, Line 21: What does LSR stand for?*

Response:
Yes, thank you for bringing that to our attention. The acronym was not defined previously, and this has been corrected in the text.

---

23. *Page 5, Line 26: Add "Herger et al, 2015" as a reference, as they also conducted a PCA and found that the global warming explains the largest percentage of variance across scenarios (see their Figure 1).*

Response:
Yes, thank you. It has been added to the sentence referenced above.

---

24. *Results/Discussion Figure 1/ Page 6, Lines 5-9: Why does Figure 1 show absolute temperature even though anomalies are used for the pattern scaling approaches? Differences between the model ensemble and reanalysis are expected to be large if one considers absolute values rather than anomalies. I recommend adding multiple observational products to this figure for validation (and replace NCEP1 with a more recent product). Moreover, I recommend terminating the time series of the observational product in 2005, as there is no meaningful comparison between observation and model data later on. Moreover, the authors could add the ensemble mean (dashed) in addition to the median (solid) to make it clearer that the shading is symmetrical about the mean.*

Response:
This figure has moved to the Supplementary Section for many reasons (see above discussion), and we agree that "there is no meaningful comparison between observation and model data later on", and as such, discussion is limited. However, this figure uses the absolute values rather than anomalies to demonstrate two things: 1. That the ensemble median and ensemble mean 1-sigma show good agreement in magnitude of year to year variation; 2. that the trend at various timespans in the historical period is not significant. We did not use anomalies in this figure because the discussion of reference epoch and difference in anomalies based upon reference epoch is better represented in Figure 2. For this figure, we also wanted to show the difference between the absolute temperature for ensemble median and ensemble mean 1-sigma for each future scenario. We would like to keep this figure as simple as possible, and as such do not think that including the model mean would be beneficial, particularly because the shading is a function of the ensemble mean.

---

25. *Figures 2-4, 6-11: To make it easier for the reader to quickly understand what is shown in the Figures, the authors should add labels (ANN, DJF, JJA, rcp, time period,...) outside the panels rather than just having it in the caption.*

Response:
Yes, thank you. We recognized that annotation of the figures is necessary, and has been done, much like you have suggested.

---

26. *Moreover, add the unit to the colorbar and rotate the colorbar labels in Figures 3, 4, 6-11.*

Response:
Yes, thank you. It has been edited with units, and color bar orientation has been adjusted.

---

27. *Figure 2: The unit should be C/year. Moreover, as the plots are symmetrical about the 1-1 axis, the authors do not necessarily need to show the lower half. They could potentially show the results of rcp8.5 in the lower left corner and the results of rcp4.5 in the upper right corner to reduce the number of subfigures from 6 to 3.*

Response:
Yes, the plots are symmetric about the 1-1 axis. We choose to keep the current plot format because it was clearer to interpret despite a reduction in the number of plots in this figure.

---

28. *Table 2: After reading Lines 17-21 (Page 6) it is not clear to me what additional information Table 2 gives compared to Figures 2 and 3, which already show the change in mean and trend. If Table 2 is kept, please add the abbreviations M21C, L19C to the caption. Moreover, the unit of the trend should be C/year rather than just C.*

Response:
We concluded that this table did not offer much more additional information, and have removed it..

---

29. *Figure 4: Add (a) to the top and (b) to the bottom plot to make a clearer distinction between those 2 plots. Then refer to them properly in the text (4a in Page 6, Line 23 and 4b in Line 26). The caption says "Ensemble mean global air temperature delta method pattern difference based on differences in historical (top) and future (bottom) reference periods". However, the reference period in both plots are within the historical period (1861-1890 or 1971-2000). Why are they referred to as historical and future?*

Response:
Thank you for pointing this out. This was a mistake. The 'reference' epochs are the late 19th and 20th Century epochs, and the future epochs are the mid and late 21st Century epochs. It has been edited to clear up this issue.
We also included additional plots that use the RCP4.5 scenario, and compare the mid 21st Century for both reference/future epochs. This has made the interpretation of epoch differences clearer.

---

30. *Figure 5: Put the months and their x-ticks in the center between the 2 boxes to make it clearer that 4 boxes belong to 1 month. Moreover, the caption does not explain what the difference between the boxes on the left and right are. Which one is the global air temperature difference and which one is the trend?*

Response:
We have edited the figure x-axis and caption to make figure clearer. We have also moved this figure to the Supplementary section in that it adds value to the storyline, but not to a significant degree.

---

31. *Page 7, Line 1: Adjust: "The different methods used to quantify global change are first examined by comparing future minus present change (left box) to the 21st Century projected linear trend (right box, Figure 5)."*

Response:
Yes, this is a confusing statement. It has been edited.

---

32. *Page 7, Line 10: Add a space between "(Figure 6)" and "and".*

Response:
Yes. Thank you. It has been edited.

---

33. *Figure 6: Add a space between "(bottom)" and "in" in the caption.*

Response:
We have edited the figure captions as well as the figure annotations.

---

34. *Figure 7: Maybe mention in the caption that the scale in the colorbar is not linear (also Figure 9). Moreover, Figure 7 could potentially be incorporated into Figure 6 with the help of stippling/hatching for significant differences. Having both the differences and significance in one Figure would make the interpretation of the results easier.*

Response:
We agree, but we have found that stippling on these plots is very difficult to read. The size and projection (we have edited the projection of the map plots) of the plots, as well as the relatively small areas where there is a significant difference (and hence stippling) have contributed to this problem.

---

35. *Page 7, Line 23: "regression pattern scaling method" rather than "regression pattern method".*

Response:
Yes, thank you. This has been edited.

---

36. *Page 7, Line 29: A high comma is missing after "drier".*

Response:
Yes, thank you. This has been edited.

---

37. *Page 8, Lines 18-19: Parentheses are for clarifications and referencing and should be avoided in this case. See, the following article by Alan Robock in EOS:*
http://onlinelibrary.wiley.com/doi/10.1029/2010EO450004/abstract

Response:
Yes, thank you. This has been edited.

Revision:
The ocean GMT sensitivity of the RCP4.5 scenario is ≥0.5^{o}C than the RCP8.5 scenario over the Arctic region, which is proportional to the land GMT sensitivity in the Antarctic.

---

38. *Acknowledgments Page 9, Line 12: The ESGF link does not seem to work (privacy issues). Please remove the "9" from the link.*

Yes, thank you. It has been edited.