

## Uncertainty-informed selection of CMIP6 Earth System Model subsets for use in multisectoral and impact models - Response to Reviewers

### Reviewer 1

The study posits a strategy for selecting 5 CMIP6 GCMs that are suitable globally for impact model applications based on temperature and precipitation characteristics and the IPCC likely ECS range.

Authors' response [blue throughout]: Thank you very much for your thoughtful review.

The results would benefit from context, both in terms of how the study compares to previous model subselection exercises (why select the same set of models for all regions?)

We have extended our motivation of the work in the Introduction and Methods Sections, noting that: "Specifically, this work is intended for use with models like GCAM, which require global coverage."

Mixing and matching ESMs across regions would result in inconsistent scenarios in that context. We have also added the clarification that if one were interested in specific regions, they could consider data from only those regions.

and in a deeper dive into the origins of the IPCC likely ECS range (where it comes from, what constraint assumptions are being made).

We have added additional discussion of the IPCC ECS range to our Methods section, as this is a critical component of the selection process.

Additionally, the methodology is hard to follow in the appendix and is worth moving to the main text.

We have moved the description of the Hawkins and Sutton values to Methods, and clarified that it is only used for independent evaluation of the model subset selection, rather than as part of the selection process. We have also clarified details of the HS calculations, as raised in points below and by other reviewers.

The primary issue I have, though, is that taking the "model uncertainty, scenario uncertainty, and interannual variability of the full CMIP6 ESM results" is inappropriate in an ensemble of opportunity like CMIP6 without a careful audit of model dependence.

This is an excellent point. This method is very flexible to different assumptions and we have added an example of incorporating considerations about model dependence, and

we have adjusted the Introduction to be clearer that this paper is describing a method for subselection rather than a Universal Best Subset.

#### Recommended Literature:

Abramowitz, G., Herger, N., Gutmann, E., Hammerling, D., Knutti, R., Leduc, M., Lorenz, R., Pincus, R., and Schmidt, G. A.: ESD Reviews: Model dependence in multi-model climate ensembles: weighting, sub-selection and out-of-sample testing, *Earth Syst. Dynam.*, 10, 91–105, <https://doi.org/10.5194/esd-10-91-2019>, 2019.

Brands, S.: A circulation-based performance atlas of the CMIP5 and 6 models for regional climate studies in the Northern Hemisphere mid-to-high latitudes, *Geosci. Model Dev.*, 15, 1375–1411, <https://doi.org/10.5194/gmd-15-1375-2022>, 2022.

Merrifield, A. L., Brunner, L., Lorenz, R., Humphrey, V., and Knutti, R.: Climate model Selection by Independence, Performance, and Spread (ClimSIPS v1.0.1) for regional applications, *Geosci. Model Dev.*, 16, 4715–4747, <https://doi.org/10.5194/gmd-16-4715-2023>, 2023.

Thank you, we have incorporated these citations into our Introduction to further highlight that the method we propose can serve as a complement to other selection criteria.

#### Specific Comments:

L60-62: "In a world unburdened by time and computing constraints, an impact model would take as input every projected data set available to have a full understanding of possible outcomes." - An ensemble of every projected data set in CMIP6 does not confer the full understanding of possible outcomes. It would include 50 initial condition ensemble members of certain ESMs and one ensemble member for others. Is the first 50x more likely to be true? Beyond the unequal voting power of the large ensembles in CMIP6, the ensemble contains a number of "hidden dependencies": models with different names but near-identical code. For uncertainty to mean "our full understanding of possible outcomes", model dependence must be handled properly.

As noted above, we have added a discussion of addressing model dependency via an additional example. To address the over-weighting of models with more ensemble members, we have also clarified the region-level indices used to characterize each model's outputs in the Methods. Briefly, the indices are calculated on each ensemble member available for a given model, but then the ensemble average of the index is the only value used for a model. In other words, models with 50 ensemble members don't get 50 entries for each index in each region. They do, however, likely have a more

confident estimate of the ensemble averaged index than say a model with only 1 ensemble member (particularly for internal variability).

L81-82: While this is an interesting objective, the size of the initial condition ensembles submitted to an exercise like CMIP is a function of computational resources and goodwill (they are submitting "free" data for others) on the side of the modeling centers. It is beneficial to many researchers that CMIP is inclusive and does not "pick favorites" thus encouraging participation.

This is a great point and it was not our intention to suggest favorites should be picked. We have adjusted this sentence. We have also removed our suggestions to CMIP modelers to leverage this idea to more efficiently deploy total computational resources. While this is technically possible to take from our work, it's also not very realistic to the CMIP process for the reasons you highlight.

Table 1: Of the 22 models you are using, 6 are connected, either by legacy or because they use a version, to NCAR's Community Atmosphere Model (CAM) development cycle. As this leaves the potential for CAM to have more influence on your uncertainty benchmark, the choice must be discussed. Additional similar models, such as ACCESS-CM2 / UKESM1-0-LL and MPI-ESM1-2- HR / MPI-ESM1-2- LR, present could be creating a "rather heterogeneous, clustered distribution, with families of closely related models lying close together but with significant voids in-between model clusters" (description of CMIP5 from Sanderson, B. M., Knutti, R., and Caldwell, P.: A representative democracy to reduce interdependency in a multimodel ensemble, *J. Climate*, 28, 5171–5194, <https://doi.org/10.1175/JCLI-D-14-00362.1>, 2015.). Could multiple similar models elevate outliers (e.g., MIROC) in your metric?

The metric is actually intended to capture outliers, although MIROC is not an outlier in this group. If we only had one representative model from each dependent group, it's possible the eigenvectors would change enough that MIROC would become an outlier in that projection space and therefore the metric would select in that case.

L166: Please justify the citation of Scafetta, 2022. See <https://agupubs.onlinelibrary.wiley.com/doi/10.1029/2022GL102530>

Thank you for bringing this to our attention. We only used this to source ECS values for two models (CMCC-CM2-SR5 and NorESM2-MM). We have removed these values and references to Scafetta from the manuscript. Like other models we could not track down ECS values for (e.g. EC-Earth3-Veg-LR), these models are included as part of the full space of model behaviors we characterize with PCA, but they are not eligible for selection to the subset.

In step 3 of Table 2, how are you computing an ensemble average for the models that only provide a single run?

The ensemble average is the average of that one realization, ie that realization's value. We have clarified this in the text.

L224: Omit "over time?"

Thank you, done.

L315: "Models who more closely match the trend of observational data (W5E5v2.0 (Lange et al., 2021)) over the historic period will have their observations hold more weight. " Why? Trends are highly sensitive to internal variability, which is inherently random in temporal phase, i.e. no reason a model and observation should have the same sequence of it. A match in trend between observations and a model over a particular time period often occurs by chance and is not indicative model performance.

Deser, C., Phillips, A., Alexander, M. A., and Smoliak, B. V.: Projecting North American climate over the next 50 years: Uncertainty due to internal variability, *J. Climate*, 27, 2271–2296, <https://doi.org/10.1175/JCLI-D-13-00451.1>, 2014.

This is an excellent point. We initially used the weighting according to matching observational data only in the Hawkins and Sutton calculations we perform to evaluate the success of our subset, not in actually selecting the subset itself. We did so following the inclusion of similar weighting in the original HS papers; the authors however do note that they did not expect uniform weights to dramatically change the breakdown of total variance. Also, we do not see this method as relating to model validation and reliability (which hard to characterize for global scale applications of multiple variables), but simply as a way to subset the large ensemble while preserving uncertainty characteristics. In this spirit, we move to uniform weights in all of our HS calculations for evaluation.

## Reviewer 2

Review of “Uncertainty-informed selection of CMIP6 Earth System Model subsets for use in multisectoral and impact models” by Snyder et al.

The presented study develops a model selection procedure that aims to preserve the distribution of total uncertainty as defined by Hawkins and Sutton. The motivation and need for some kind sub-selection of global climate models for impact models is clear and well-motivated by the authors. Their method is fairly straight-forward and is shown to work for the metric the authors use as validation.

I think this approach could be valuable contribution to be used as an objective selection criterion if the authors manage to manage to better describe the caveats and limits of this approach and better showcase the effect of the sub-selection. In its current state I can not recommend the manuscript for publication based on the major comments outlined in the following.

Thank you, we respond to each comment below, in blue.

### Major comments

I am very critical about the idea of optimizing a sub-selection method only for the distribution of the three sources of uncertainty. I note that the authors also manually select by ECS but this seems to be not really part of the method and is presented more as an afterthought.

Thank you, we have edited the text to more heavily emphasize the role of ECS in the Methods, which was always a critical aspect of our approach but evidently was not highlighted sufficiently in its importance.

In any case, such an approach opens up the risk of selecting highly dependent models (as seems in fact to be the case in the presented example with both ACCESS versions being part of the suggested sub-ensemble) as well as objectively bad models. Both are situations which should be avoided I would argue. In fact, in a recent study Merrifield et al. (2023; 10.5194/gmd-16-4715-2023) set out to solve a similar problem but also consider model performance and independence. Their results should at least be compared to the results presented in this manuscript.

Thank you - we have added a citation to Merrifield et al as well as a second experiment to the manuscript to highlight the flexibility of the method. In the second experiment, we consider an extremely simple definition of independent models to focus on, and repeat the selection process. As you are aware, there are a wide variety of definitions and criteria to define model independence, and the question is not yet settled in the

literature. We hope that while our definition of independence in our added Experiment 2 is simple, it highlights that the method *could* be applied and validated for a different set of independent models making up the 'full data' based on a different definition of independence. In other words, we think our method can be a powerful complement to other methods concerned more heavily with independence. At the very least, a researcher performing this experiment would likely learn something about their definition of independence.

I have to admit I do not fully understand the role of the scenario uncertainty this manuscript. In an idealized case the scenario uncertainty as defined by HS is only dependent on the scenarios used. Why is it considered here, what is its relation to model selection? If it is sensitive to the selected (ECS preserving) model ensemble this is only a sign that the HS method is not properly able to isolate forced response, model uncertainty, and internal variability (as discussed in Lehner et al. 2020) is it not?

Thank you for highlighting this. It is true that we are looking to represent model uncertainty which is reflected in both T&P anomalies and IASD. But we also wish to do so in a robust way, and for the CMIP6 ensemble that means we need to consider the scenarios that were run (at least the Tier1 scenarios, here). We aren't sub-selecting for scenarios but the scenario space is vital to appreciate the variability of the different models. We have attempted to clarify this in the methods section describing the HS fractions and what we use them for.

With respect to the verification metric, I am not sure if optimizing for the relative distribution of uncertainty is really ideal. At least the authors should consider also looking at the absolute position of their selected models in temperature-precipitation change space. I could imagine a situation where the relative distribution of uncertainties in the subset is very similar to the full ensemble without the subset being representative when considering, e.g., absolute changes.

We have moved the description of the Hawkins and Sutton values to Methods, and clarified that it is only used for independent evaluation of the model subset selection, rather than as part of the selection process. We have further emphasized that the selection process only takes place in the absolute position/changes space of T and P anomalies/IASD and not the relative space of the HS fractions.

Minor comments

line 33: “Scenario simulations from CMIP (most recently through ScenarioMIP, (O’Neill et al. 2016)”

closing braked missing

Thank you, fixed.

36: “Using such multi-model ensembles captures the process and structural uncertainties represented by sampling across ESMs, scenario uncertainty,”

Minor point but a multi model is not necessary to represent scenario uncertainty?

Similar statement also in the abstract: “In this work, we present a method to select a subset of the latest phase, CMIP6, models for use as inputs to a sectoral impact or multisectoral models, while still representing the range of model uncertainty, scenario uncertainty, and interannual variability of the full CMIP6 ESM results”

Thank you for highlighting this. It is true that we are looking to represent model uncertainty which is reflected in both T&P anomalies and IASD. But we also wish to do so in a robust way, and for the CMIP6 ensemble that means we need to consider the scenarios that were run (at least the Tier1 scenarios, here). We aren’t sub-selecting for scenarios but the scenario space is vital to appreciate the variability of the different models. We have attempted to clarify this in the methods section describing the HS fractions and what we use them for.

45: “For Earth system modelers, variability across ESMs’ projections of future climate variables can be significant (Hawkins and Sutton 2009; Hawkins and Sutton 2011; Lehner et al. 2020) and so the participation of multiple modeling centers running multiple scenarios is critical to understanding the future of the Earth system.”

I agree with this statement but “variability across ESMs” is a bit vague. Of the three sources of variability considered in the cited studies only one really needs a multi-model ensemble. I am also not sure if these are really the right studies to cite here as they mainly look at relative contributions of individual sources of uncertainty.

We don’t entirely understand what the reviewer means with this comment. However, we do not think it is possible to robustly assess uncertainty in internal variability or scenario without a multi-model ensemble. We are happy to respond to and address any clarifications on this comment.

57: “However the total burden across modeling centers to sample across ESMs and scenarios still remains high, even with this potential efficiency.” I am not sure I understand the first part of this sentence?

Even if emulators can lighten the computational effort in scenario space, modeling centers are still addressing competing demands from different experiments’ goals. We have attempted to clarify in the introduction.

72: “these competing priorities” what are the ‘competing priorities’ here?

We have attempted to clarify this text.

Tab 1: I am a bit confused about the use of the word ESM in this manuscript. This table clearly also contains models versions with are not ESMs but GCMs? (ACCESS-CM2 for example)

Thank you for pointing this out, we have corrected the language throughout the manuscript to refer to the collection of models simply as ‘the models’ when a shorthand is desired, and we have clarified that both ESMs and GCMs are considered. We have also reminded readers that even Earth System Models are run in concentration-driven mode rather than emissions-driven mode for these experiments in CMIP6.

108: “and the interannual standard deviation”

It is unclear to me how this is done or in which time-period.

We have clarified this in the text: “Interannual standard deviation is calculated by detrending the regional average level temperature and precipitation time series from 1994-2100 using non-parametric locally weighted smoothing (LOESS as implemented in the python statsmodels package), and then taking the standard deviation of the residuals.

114: At this point I am also wondering how initial-condition members are considered exactly? It seems that the indices would be sensible to the number of ensemble members?

We have clarified our description of the indices in the text: “For each scenario, region, and available ensemble member in each climate model we extract the following



temperature and precipitation outputs: mid-century (2040-2059) average anomaly relative to that model's historical average (1995-2014), the end of century (2080-2099) anomaly relative to historical average, and the interannual standard deviation. Interannual standard deviation is calculated by detrending the regional average level temperature and precipitation time series from 1994-2100 using non-parametric locally weighted smoothing (LOESS as implemented in the python statsmodels package), and then taking the standard deviation of the residuals. For each scenario and model, these ensemble-member values are used to calculate the ensemble average to form our final indices in each region."

Please also see our response to R1 in regards to a similar quest: "Briefly, the indices are calculated on each ensemble member available for a given model, but then the ensemble average of the index is the only value used for a model. In other words, models with 50 ensemble members don't get 50 entries for each index in each region. They do, however, likely have a more confident estimate of the ensemble averaged index than say a model with only 1 ensemble member (particularly for internal variability)."

Fig 1: this figure seems unproportionally big for the amount of information it conveys.

Thank you for this suggestion, we have decreased the size to accommodate the scree plots for both experiments we now discuss in the text, and (briefly) further explained why we check the scree plots to determine how many eigenvectors to retain.

Fig 3: I am not quite sure what information this figure is supposed to convey to the reader? In particular, since it is hardly discussed in the text. For example, is it somehow relevant that MPI or MRI (hard to distinguish) seem to behave quite different in PC2?

Thank you - we have added more motivating text to where this figure is presented and throughout the methods.

165: "Core Writing Team & (eds.)," strange citation

We have adjusted the citation in the bib file so the editors' names are included but the IPCC requested citation is indeed a little odd: "IPCC, 2023: Summary for Policymakers.

In: Climate Change 2023: Synthesis Report. Contribution of Working Groups I, II and III to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change [Core Writing Team, H. Lee and J. Romero (eds.)]. IPCC, Geneva, Switzerland, pp. 1-34, doi: 10.59327/IPCC/AR6-9789291691647.001"

How many subsets fulfill this criterion?

We have added this number - out of 22 choose 5 potential subsets, there are 72 subsets available meeting this constraint using all models for which we could identify reliable ECS values.

Fig 4: Again I am not sure what I learn from this plot.

Thank you, we have expanded our discussion of this plot.

199: "To manage the inspection of three time series" Is 'three time-series' referring to Hawkins and Suttons model, forced, and internal components? This is not clear to me here. If yes I would note that in HS internal variability is not a timeseries but constant over time if memory serves.

We have clarified that the three time series is referring to the fraction of total uncertainty described by the model, forced, and internal components. The HS internal variability is itself a single value, but that single value makes up a different fraction of the total uncertainty as the model and forced components vary over time. We have clarified in the text.

223: "For temperature, we see that interannual variability is often performing well, with increasingly better performance over time? over time." Typo

Thank you, corrected.

### Reviewer 3

Review of esd-2023-41

Title: Uncertainty-informed selection of CMIP6 Earth System Model subsets for use in multisectoral and impact models

Authors: Abigail Snyder, Noah Prime, Claudia Tebaldi, Kalyn Dorheim

*Overall Recommendation: Minor Revisions*

This study, submitted to Earth System Dynamics, highlights a new approach to select subsets of ESMs (and potentially GCMs) for use in multi-sectoral and impacts modeling. The approach is novel and attempts to capture multiple sources of uncertainty in the subset and does have potential utility for multiple applications beyond what was mentioned in the manuscript. However, I do think the manuscript could be improved by added justification in a few places and addressing some particular aspects from prior literature that is relevant. In particular, the attempt at subset selection itself is not a new effort and the authors should place this manuscript in the context of prior attempts at subset selection. My general comments to improve the manuscript are below, followed by specific comments. I look forward to seeing the revised manuscript in print.

[Thank you, responses in blue to each comment below.](#)

#### General Comments

There are a number of current issues in literature related to the selection of ESMs (or even just a set of climate projections) for use in applications that were not discussed here but are quite relevant.

First and foremost, the challenge of ensemble subset selection is not new. It would be worthwhile given the focus on using ESMs in impacts models to place this article in context with existing literature. For example, Parding et al (2020) produced the GCMeval tool, which is designed as an interactive tool for evaluation and selection of climate model ensembles for use in multiple applications. The methods in the GCMeval tool are relatively simplistic. In the context of this manuscript, I am left to wonder how different this approach is to the approach of Parding et al (2020) or other previous research. It would strengthen the manuscript to briefly discuss the differences between this approach and others in prior literature. Ideally, there should be some analytic

comparisons between these approaches, but I believe a brief discussion of the other literature on this topic would suffice for this manuscript. In addition, the authors are addressing the “practitioner’s dilemma” (Barsugli et al. 2013). Though it was not mentioned by name, it may also be worthwhile to discuss how this approach builds on the literature associated with this well-known challenge.

Thank you for these recommendations - we have added further discussion of these in our introduction and throughout the manuscript. We have clarified the framing of our manuscript as being an uncertainty-first focused method that can be used potentially in complement to methods for evaluation and model dependence such as those in Parding et al, Merrifield et al, and others. We have also clarified in our framing that the primary use-case in mind for this method is for models that require global coverage of climate data (such as GCAM and similar models) rather than more local phenomenon. Discussion of downscaling and the practitioner’s dilemma has been expanded in our conclusions.

Second, the approach in this manuscript may well retain “hot-models”, particularly as one goes to large subset sizes. The literature regarding if one should use hot-models is mixed, ranging from omitting hot models entirely (Hausfather et al. 2022), to down-weighting hot models based on ECS (Massoud et al. 2023), or simply keeping hot models as they may not have serious impacts on impacts modeling (Asenjan et al. 2023). It would be worthwhile for the authors to address how their method handles the “hot models” problem specifically, when their approach may indeed retain these models in their subsets.

Thank you for this note. In response to this and comments from the other reviewers, we have more strongly emphasized the role of constraining our model subset to match the IPCC most likely distribution of ECS values. This of course allows one high ECS model to be selected in the subset to explore potential ‘worst case’ impacts downstream, but overall moves the ‘hotness’ of our subset to be lower than the full collection of voluntarily submitted models. In addition to emphasizing this more clearly in the methods, we incorporate this into the discussion of our results as well.

Third, the focus is on impacts modeling with ESMs, but this is not the only use for ESM output. Depending on the needs of a stakeholder, one may not need additional modeling, but rather require complex variables derived from ESM output. A prime example of this is the use of spring phenological indices to determine projected changes in first leaf, first bloom, or the likelihood of a false spring (Gerst et al. 2020; Allstadt et al;

2015; Peterson and Abatzoglou, 2014). The approach in this manuscript could well be used in situations that don't require additional modeling per se, but do require derivation from ESM or downscaled ESM output. Such literature should be briefly mentioned with the authors comment.

Thank you, we have added a discussion of considering more derived variables to our conclusions.

Fourth, the Hawkins and Sutton (2009) approach does have some valid criticisms. I suggest looking at the recent work of Lafferty and Sriver (2023), which also uses Hawkins and Sutton (2009) and the work of Wootten et al. (2017). The Lafferty and Sriver (2023) article addresses the critiques around the Hawkins and Sutton (2009) approach.

While we agree that no method is perfect, the Hawkins and Sutton approach remains a useful framework for the qualitative evaluation of how well the selected subsets of climate models preserve uncertainty characteristics of the full data. We find it especially useful for understanding whether discrepancies in breakdown of total variance for the full vs subset data are explainable by our choice to constrain ECS distribution as part of the selection process as a way to address the hot model problem. We have adjusted the text to emphasize these points, and that the Hawkins and Sutton breakdown are not an active part of the selection procedure but rather a post-hoc evaluation tool, throughout the text. We have also clarified in the introduction and conclusions that this method is meant more for studies using multisector dynamic models that require global coverage, where consistency in handling regions is key. We have also added citation to Lafferty and Sriver in our discussion of how this work interacts with bias correction and downscaling considerations in the conclusion, in response to this and the next comment.

Finally, while the selection of ESMs is important, most impacts modelers do not use the ESMs directly, but use the downscaled ESM output (whether dynamically or statistically downscaled). The authors mention this briefly in passing, but it is important to acknowledge this in the conclusions also. Downscaling is itself another source of uncertainty, so it is a question of if this approach could also be applied with ESMs downscaled with multiple approaches.

We have added a brief paragraph to the conclusions regarding bias correction in downscaling. In general, we do not recommend applying this method be applied to already-downscaled data that has been downscaled by different methods. While bias-correction and downscaling is expensive, we believe the consistency gained from using a single robust method for all climate data sets is worth this. Indeed, one

motivation for us as users of this method is to only have to bias-correct and downscale the output of 5 models instead of 22 using the ISIMIP3b protocol.

Literature mentioned above:

Allstadt, A. J., S. J. Vavrus, P. J. Heglund, A. M. Pidgeon, W. E. Thogmartin, and V. C. Radeloff, 2015: Spring plant phenology and false springs in the conterminous US during the 21st century. *Environmental Research Letters*, 10, <https://doi.org/10.1088/1748-9326/10/10/104008>

Asenjan, M.R., F. Brissette, J.-L. Martel, and R. Arsenault, 2023: Understanding the influence of “hot” models in climate impact studies: a hydrological perspective. *Hydrology and Earth System Sciences*, 27, 4355-4367, DOI: 10.5194/hess-27-4355-2023

Barsugli, J., and Coauthors, 2013: The Practitioner’s Dilemma: How to Assess the Credibility of Downscaled Climate Projections. *Eos Transactions*, 94, 424–425, <https://doi.org/10.1002/2013EO460005>.

Gerst, K. L., T. M. Crimmins, E. E. Posthumus, A. H. Rosemartin, and M. D. Schwartz, 2020: How well do the spring indices predict phenological activity across plant species? *Int J Biometeorol*, 64, 889–901, <https://doi.org/10.1007/s00484-020-01879-z>.

Hausfather, Z., K. Marvel, G. A. Schmidt, J. W. Nielsen-Gammon, and M. Zelinka, 2022: Climate simulations: recognize the ‘hot model’ problem. *Nature*, 605, 26–29, <https://doi.org/10.1038/d41586-022-01192-2>.

Lafferty, D.C. and R.L. Sriver, 2023: Downscaling and bias-correction contribute considerable uncertainty to local climate projections in CMIP6. *Nature Partner Journal – Climate and Atmospheric Science*, 6, doi: <https://doi.org/10.1038/s41612-023-00486-0>

Parding, K. M., and Coauthors, 2020: GCMeval – An interactive tool for evaluation and selection of climate model ensembles. *Climate Services*, 18, 100167, <https://doi.org/10.1016/j.cliser.2020.100167>.

Peterson, A. G., and J. T. Abatzoglou, 2014: Observed changes in false springs over the contiguous United States. *Geophysical Research Letters*, 41, 3307–3314, <https://doi.org/10.1002/2014GL061184>.Received.

Wootten, A., A. Terando, B. J. Reich, R. P. Boyles, and F. Semazzi, 2017: Characterizing Sources of Uncertainty from Global Climate Models and Downscaling

Techniques. *Journal of Applied Meteorology and Climatology*, 56, 3245–3262, <https://doi.org/10.1175/JAMC-D-17-0087.1>.

Minor Comments:

Line 84, first mention of Table 2: Table 2 defines the process for selection to aid the reader, yet it is positioned in the text far from the section. It's also mentioned before Table 1 and Table 1 appears sooner in the text. I suggest reordering Table 2 and Table 1 and placing the renamed Table 2 earlier in the text to help the reader.

[Thank you, we have done this.](#)

Line 105: “For each scenario and region in each ESM,...” – Am I correct that the averages calculated are across all the initializations of each ESM? Or is it the average across all models?

[Yes, the average is across all initializations of each model within each scenarios, not across models. We have clarified this in the revised methods.](#)

Lines 127-128: “Based on this figure,...explaining 71.8% of variance.” – Why only 5 eigenvectors? Why not more?

[Thank you, we have added text explaining that we cut eigenvectors just after the major ‘elbow’ in the scree plot, which is a pretty common rule of thumb albeit there are no hard and fast rules. There’s no reason this method couldn’t be applied on more, or all, eigenvectors. But for the sake of faster evaluation and easier visualization, we stop at 5.](#)

Line 224: “...better performance over t ime? over time.” – This is a typo.

[Thank you, we have corrected this.](#)

Lines 220-240: It seems the discussion focused on temperature plots only, but I didn’t see any comparison of temperature vs. precipitation results.

[Thank you, we have added some discussion of this.](#)