

## **Authors' response to Anonymous Referee #1**

*The authors present a review of a wide array of studies that address model dependence in multi-model climate ensembles. They also provide valuable guidance on how users can test the efficacy of approaches that account for dependence. This is a very timely contribution coming as it does in the months leading up to CMIP6 –easily the largest climate modeling exercise ever carried out. This manuscript will set the stage for how model simulations are treated – and hopefully move beyond equally weighted simulations.*

We are pleased that the reviewer feels this work is useful and timely, and thank them for their efforts. We endeavour to address their suggested improvements below.

*My main comment is that the manuscript needs to be better organized. It currently has 11 sections and no roadmap up front on what these sections contain. Surely, there is a way to consolidate some of these into subsections and have fewer main sections. Most of the comments below are also related to readability.*

Yes, we agree that a roadmap in the earlier stages of the manuscript would be helpful, and we will work to implement one. Regarding consolidation, we have gone backwards and forwards on this issue. The extensive subsection approach came after asking colleagues to suggest improvements to this manuscript, and we feel that it is more readable with this additional discretisation. If the editor feels strongly about this we can of course revert to using fewer subheadings but we genuinely feel it is better as is.

**Proposed revision:** introduce a roadmap of sections in Section 1, and work to structure narrative in terms of this roadmap.

*1. Page 2, Line 8: The six sources of uncertainty are in the first paragraph.*

Yes, we will amend this to refer to the first instead of second paragraph.

**Proposed revision:** change text to “six sources of uncertainty affecting this likelihood in the first paragraph”

*2. Page 2, Line 25: The word “calibrated” may be replaced with “constructed” or “designed”.*

We are happy to change to one of these.

**Proposed revision:** replace “calibrated” with ‘constructed’.

*3. Page 2, Line 36: Again the word calibrated is used in a different sense than is standard.*

We have no objection to using ‘constructed’ in this context instead.

**Proposed revision:** replace “calibrated” with ‘constructed’.

*4. Page 4, Line 6 and Line 29: The analogies from evolutionary classifications may not help the typical reader or even make a connection. I suspect many, like this reviewer will have to look it up.*

We agree that not all readers will necessarily have the background, but feel like this analogy does add to the paper - most people have thought about the process of evolution in a biological context, at least to some degree.

**Proposed revision:** The sentence will be reorganized to present more context to the reader before the analogy.

*5. Page 4, Line 20: The use of “first version number” and “second version number” are actually better understood as MajorRevisionNumber and MinorRevisionNumber. Boé (2018) provides a version*

number example that avoids the confusion in this manuscript. Perhaps move the CLM4/CLM4.5 example up to avoid this confusion?

Yes, agreed.

*Proposed revision:* We will move the CLM example up and replace existing text with the Major/Minor revision number delineation.

6. Page 4, Line 24-26: “A reminder . . .” is a complicated sentence. Perhaps rewrite in simpler language.

Yes, agreed.

*Proposed revision:* “As discussed above, an ideal definition of model dependence would only include variability in process representations that are not tightly observationally constrained (so that several models using the Navier-Stokes equations might not represent dependent treatment of process, for example).”

7. Page 5, Line 32: In the sentence “. . .these approaches could in principle address many of the shortcomings of approaches such as those above. . .” I believe the latter “approaches” refers to the 3 discussed in the preceding paragraphs. This sentence should be reworded to make it clear.

Yes, correct. We agree this should be changed.

*Proposed revision:* “. . .these approaches could in principle address many of the shortcomings of relatively simplistic techniques, such as those in the paragraphs above. . .”

8. Page 6, Lines 1-7: Is there an implicit assumption (in figure 1) that observational estimates are close together? In the regional context, it is quite common for multiple observational estimates to be further apart than inter-model distances.

Not necessarily, no, this is really just for illustration. It seems reasonable to assert that with large observational uncertainty relative to ensemble spread, making categorical statements about model dependence is much more difficult.

*Proposed revision:* add text to make it clear that the observational estimates needn't be close together when describing Figure 1.

9. Page 8, Line 28: The word “calibration” appears for the first time in the section heading. Some prior context is important.

Yes. There are obviously several ways to address this issue. We propose renaming the section.

*Proposed revision:* Rename this section as “Robust strategies for addressing model dependence”

10. Page 8, Line 30: “... how might proposals to account for independence interpreted?” is missing a “be”.

Yes, correct.

*Proposed revision:* “... how might proposals to account for independence be interpreted?”

11. Page 9, Line 6: “data” here refers to “observed data” right? This needs to be clarified.

Yes, correct again.

*Proposed revision:* replace “data” with “observed data”.

12. Page 10, Lines 40-41: The sentence “Dependence is not a property of a model simulation per se, rather a property of a specific quantity in a particular simulation with respect to the rest of an ensemble” is a variant of the immediately preceding sentence.

We disagree. The first sentence notes that an assessment of dependence is sensitive to the cost function, variable, time period, location, observational product, and more. The second sentence notes that even for a given [cost function, variable, time period, location, observational product etc], a particular ensemble member does not possess dependence or independence (as it would performance) except in relation to the other members of the ensemble. That is, the same ensemble member might be highly dependent, or highly independent, depending on the makeup of the remaining ensemble. We propose to amend this text to make the distinction clearer.

*Proposed revision:* “The sensitivity of weighting and sub-ensemble selection results to metric, variable, observational estimate, location, time and spatial scale and calibration time period underscores that a characterisation of model dependence is not a general property of an ensemble, but application-specific. Also, note that dependence is not a property of a model simulation per se, as performance is, but rather a property of the simulation with respect to the rest of the ensemble.”