

Dear Author,

Thank you for the revised version of your manuscript. I have to say that I had to spend a significant amount of time to analyze your responses to Referee #2's comments. And I take the opportunity to thank Referee #2 for his detailed comments that certainly took him an even more significant amount of time to write. As Referee #2, I don't criticize the content of the paper but its form. With this new version, the manuscript has improved but what Referee #2 wrote is still true:

- some paragraphs contain parts that remain unclear or vague, and some sections are too long;
- ideas should be more precise, without multiple sentences containing very similar messages or information; repetition should be avoided;
- the reasoning within a paragraph should be clear, logical and well-structured; and
- the style should be more homogeneous throughout the text.

In many places, the sentences are over convoluted and too many (useless) words are included. Also, in many places, useless sentences or parts of sentences could be removed as they contain obvious ideas and do not bring any addition to the text. Next time, I strongly suggest to the author to have the manuscript reviewed by a colleague to simplify the text and make it more concise and right to the point.

With this in mind, here are additional simplifications that I propose, in addition to the ones that you already included in response to Referee #2's comments (the line numbers refer to version 3 of the manuscript). I also make additional remarks here below.

#### Additional propositions of simplification

L. 57-61: "The climate modeling community creates multi-model ensembles of a common set of ESM experiments with the same perturbations applied. The simulated climate responses can differ across a multi-model ensemble. This diversity in responses may for instance be due to differences in process complexity and interactions within the respective ESMs. Experimental protocols are used to create multi-model ensemble simulations for specific ESM experimental setups. These aim to better understand the reasons for the diversity in climate responses and feedback and to create future climate projections." : First and before last sentences are redundant; I suggest: "The climate modeling community creates experimental protocols to set up multi-model ensembles of a common set of ESM experiments. The simulated climate responses can differ across the multi-model ensemble members in response, for instance, to differences in process complexity and interactions within the respective ESMs. The aim is to better understand the reasons for the diversity in climate responses and feedback."

L.87: "... consider structural differences concerning the design and the level of complexity between ESMs." => "... consider structural differences in the design and complexity of the different ESMs."

L. 90: "... radiative forcing diversity to anthropogenic perturbations ..." => "radiative forcing linked to anthropogenic perturbations ..."

L. 173: "The RFMIP protocol included experiments to diagnose radiative forcing for greenhouse gases and aerosols as bulk quantities with setups parallel to the CMIP6 experiments for the "Diagnostic, Evaluation, and Characterization of Klima" (DECK)" => "The RFMIP protocol included experiments to diagnose radiative forcing from greenhouse gases and aerosols as bulk quantities with setups common to CMIP6 "Diagnostic, Evaluation, and Characterization of Klima" (DECK) experiments."

L. 216-220: "A major challenge to further advancing the understanding of climate change with ESMs is that differences in their results for individual segments of the perturbation-response paradigm are not independent of other segments. Specifically, a model-to-model difference in a climate response might be caused by various segments in the paradigm. For instance, the same emissions can lead to different ERFs, the same ERF can induce different climate responses and the same response can trigger different feedbacks across multi-model ensembles. In multi-model studies, one therefore sees inter-model spreads in forcing for the same change in atmospheric composition and model-dependent climate responses to the same forcing involving different types and magnitudes of feedbacks."

The text you added makes the paragraph even more fuzzy from my point of view. In the first sentence, you discuss dependence of what you call the “segments of the perturbation-response paradigm”. In the second sentence, you discuss the fact that differences may be linked to more than one segments; I am not sure then why the second sentence starts with “Specifically”. Then, if I understand well, in the 3rd sentence, you give an example of different responses in different segments. Please simplify. Consider removing the second sentence. In all cases, remove “across multi-model ensembles” in “different feedbacks across multi-model ensembles” as I think you mean that the difference in the feedbacks are differences across the ensemble members (but this is confusing as one may understand that the feedbacks themselves occur between the ensemble members).

L.226: “for a better understanding of reasons for model differences” => “for a better understanding of model differences”

L.249: “newly requested diagnostic output that is not yet available in the standard variable list of ESMs, e.g., for RFMIP-IRF” => remove “that is not yet available in the standard variable list of ESMs”

L.252-255: “In this case, it takes longer to finish the experiments and to do the associated scientific exploitation, e.g., in the case of RFMIP-SpAer several years after the work began (Fiedler et al., 2023), which is long compared to easy experiments that modelers can quickly set up via a simple change in a setting for performing an experiment, e.g., for RFMIP-ERF, thanks to prior work on the development and testing of models.” This is a way too long and over convoluted sentence and does not bring much to the previous sentence. I suggest removing it and just modifying the end of the previous sentence as “...to carry out the work including coding, testing, performing the experiments and associated scientific exploitation.”

L.265: “we reviewed the current status of the experiments” => “we reviewed the status of the experiments”

L.287: “They contribute to the decision for which community-driven MIP experiments with the ESM will be conducted, e.g., through granting computational resources and prioritizing experiments to be completed.” => “They contribute to the decision as to which MIPs will be conducted with which priority, e.g., through granting computational resources.”

L.288-290: “Additional decisions for the experiments are made by the scientists interested in the MIP. There is some room to make their own choices since not all experimental settings are explicitly defined by the MIP’s experiment protocols, e.g., they may use a coarser spatial resolution and to some degree less model complexity to reduce the computational burden.”

Another example of the same idea expressed twice. Simply remove this paragraph as the same ideas are expressed with slightly different word in the paragraph just after.

L.291: “There are inevitable tradeoffs in the final experimental designs for individual MIPs. Such choices can be categorized along the three axes of (1) *model complexity* addressing how many process interactions ESMs allow or how much fidelity processes have, (2) *model resolution* referring to the grid spacing of the model, and (3) *simulation length* covering the length and number of simulations in an ensemble of different experimental setups per ESM.” => “There are inevitable tradeoffs in the final experimental choices that can be categorized along the three axes of (1) *model complexity* addressing how many process interactions ESMs allow or how much fidelity processes have, (2) *model resolution* referring to the grid spacing, and (3) *simulation length* covering the length and number of members in ensemble simulations.”

L.339: remove “of the experiments” in “the ensemble sizes and simulation lengths of the experiments”

L.445-455 (L. 427-434 in version 2) : Referee #2 wrote that these sentence are “vague, not nice” but you did not change anything. Please rephrase, clarify, simplify.

L.528-530 : “Much finer spatial resolutions with horizontal grid spacings of a few kilometers hold the potential to overcome some of the long-standing challenges concerning the representation of clouds, precipitation, and circulation in global climate simulations, which would require a step change in collaboration between climate science and high-performance computing (Slingo et al., 2022).”

is another example of a too long sentence; please cut in two parts, end first sentence after “global climate simulations”; in the second part, it is not clear what “which” refer to, please rephrase.

L.531 : Change “in coarse resolution models of several tens to hundreds of kilometers of grid spacings” for “in models with resolution of several tens to hundreds of kilometers”.

L. 560-561: “especially in the context of a MIP since fully coupled ESMs with interactive aerosols and chemistry at a resolution of 1 km fast enough to perform multi-decadal simulations are unlikely to be ready in the time of CMIP7”: please rephrase for something like “since fully coupled kilometer-scale ESMs with interactive aerosols and chemistry fast enough to perform multi-decadal simulations are unlikely to be ready in the time of CMIP7”

L. 564-566: “This approach is suitable to answer some but not all research questions in our community. For instance, the response of dust emission fluxes to changes in winds and moisture can be addressed with offline modeling and allows to identify underlying reasons for changes and model differences in the dust response (Fiedler et al., 2016), but the implication of such dust emission changes for air quality and climate responses can not be quantified with such an approach.” ; please simplify for something like “This approach is suitable to answer some but not all research questions in our community, for instance, the response of dust emission fluxes to changes in winds and moisture (Fiedler et al., 2016), but not the implication of such dust emission changes for air quality and climate responses.”

L. 691: “sufficiently long experiments or sufficiently many ensemble members” is not grammatically correct; I suggest “sufficiently long experiments or enough ensemble members”

L 692: “to better distinguish climate and air quality responses to atmospheric composition changes from” is another example of useless words in a sentence; it should simply be “to better distinguish climate and air quality responses to atmospheric composition from”

L.711: “multi-purpose experiments can be useful and less burden some in terms of human and computational resources” => “multi-purpose experiments can be useful and associated to less burden in terms of human and computational resources“

L.712: “as long as they facilitate answering the science questions laid out by the MIPs” is one example of an obvious sentence that does not bring much to the text

Figure 1 captions: “are simulated with a model configuration coupled to an ocean model.” => “can be simulated with a model configuration including coupling to an ocean model.”

#### Additional comments and propositions of modifications

L.62-64: “Results from multi-model intercomparison projects (MIPs) are widely used to advance scientific understanding and inform stakeholders on climate change. The most prominent example is the Coupled Model Intercomparison Project (CMIP, Meehl et al., 2000) that has contributed ...”: I don’t agree that CMIP is an example of à MIP; instead, CMIP is composed of several MIPs, please rephrase.

L.85-100: Some lines use the past tense and other the present tense, please standardise .

L.88: “for the answer to the MIP’s question” => “in the answer to the MIP’s question “

L.115: “and refers to the same term as short-lived climate forcers (SLCFs) used by Collins et al.” => “and refers to the same concept than short-lived climate forcers (SLCFs) used by Collins et al. “

L.155: I don't understand what "quantify radiative forcing within and outside of the three MIPs" means, please clarify.

L.268: Change "out of the" for "of the" in "Available model output to assess differences in forcing and response was, however, limited, e.g., output for the mid-visible aerosol optical depth is available only for 45 out of the 67 models providing *historical* experiments."

L.345: "and also pose challenges for reducing model-based uncertainty" => "but pose challenges in reducing model-based uncertainty"

L.365: remove "in" in "do not prescribe the level of process complexity in a ..."

L.545: "globally for restricted time periods of a few weeks to years" : be more specific on the number of years so to be coherent with the first part of the sentence.

L.569: make a new paragraph before "We perceive dynamical downscaling ..."

L.645: As suggested by referee #2: "There is the opportunity" => "There is an opportunity" (you did not make the change).

L.710: "is smaller for the separate MIPs" => "than what it would be for one converged MIP"

Table 1: The additions made in the legend are not self-coherent. For example "hist-X" does not appear as is in the column (does it refer to "hist" and/or "hist-piAer" ?) while "histSST-X", "piClim-X and "piClim-2xX" do appear as is. Also "piControl" and all "ssp370..." are not described; please change.

Finally, I see that Table 4 has thoroughly changed without justification; can you make some comments on why you made all those changes?