Sungbo Shim (CC1):

According to line 242-250, it is stated that CMIP will provide CO2 concentration data for GCM without carbon cycle module. Do you mean that CMIP will provide concentration data for all 6 future scenarios (Tier1) in Table 1?? Or do CMIP only provide the concentration required for the 3 experimental types (HC, MC, LC) in Table 1?? These sentences are not clear. I would like to express it in detail.

Yes, concentration data will be provided for all scenarios. We have clarified in the text.:

Jean-Francois Lamarque (RC1):

This paper discusses a proposal for the upcoming CMIP7 ScenarioMIP. It provides a presentation of the main aspects driving the design for the main experiments. This paper is designed to be of use by both the IAMs and ESMs communities, which tends to make the paper somewhat confusing at times, as identified in some of my comments below. Maybe the paper would be easier to read if the discussions for the two communities were separated? Furthermore, while it is vying to provide a complete picture, I have found many instances of unclear statements and unjustified decisions. The paper will be significantly improved by considering the following comments, listed as they appear in the paper

Thank you for your thoughtful comments, we will consider them carefully. Regarding the more general point about speaking to two communities (in fact, we would argue three, including the impact research community): this is indeed a complicating factor, but that is also what makes ScenarioMIP unique. So we would like to keep that aspect, and think that the detailed comments made by you and other reviewers can help us to make the text clearer.

The title should indicate that this is a proposal, since that is made very clear in multiple places in the article

We would like to keep the current title for consistency with the CMIP6 version of the paper (O'Neill et al., 2016) that served the same purpose.

Lines 111-114: this seems like a much bigger role than what this paper describes, which is a set of scenarios. Similarly, line 199 should be “participants in ScenarioMIP”

As part of the scenarios’ forcings we plan to deliver those types of datasets, so this is within the purview of ScenarioMIP. Assuming the reference is to line 119, not 199, we are now referring to climate modeling groups participating in ScenarioMIP.

Line 137: instead of “IAM scenarios”, should it be “IAM-generated data based on the scenarios in this proposal”? Otherwise “IAM scenarios” should be defined.

We have expanded the text here to indicate that we mean the IAM emissions and land use scenarios based on the experimental design described in the paper.

Lines 163-165: this is an interesting point, although it seems that the end result is still with a high RCP8.5-like (or close) scenario. So, were the critiques unfounded and/or ignored?

Not ignored, but possibly one of the major critiques, having to do with climate impacts on the economy, remains unaddressed as methodological challenges still prevent us from representing them in the type of IAM simulations that will deliver the ScenarioMIP scenarios. That said, the results of “pushing the envelope of the upper end of emissions” results in a scenario that is far from reaching 8.5W/m^2, and is in fact even below 7W/m^2 according to FaIR preliminary estimates.

Line 195: what is the rationale for this “agnostic” position? This seems like a step back towards the RCPs

This was always the principle behind the scenario framework, i.e., the fact that different SSPs could lead to the same RCP, and different RCPs could be compatible with the socio-economic assumptions of the same SSP. In CMIP6 the choice was made to explicitly mention the driving SSP of the marker scenario in the nomenclature of the forcing pathways and *a posteriori* we found this to be misleading. The assumption is that the socio-economic details of the SSPs are responsible only for spatial patterns of land use and SLCF emissions whose signal, in the multimodel context, is not robust. So, although one specific choice of SSP will underlie the marker chosen for each scenario, the expectation is that a different choice that still achieves the same global level of

emissions of well mixed greenhouse gases would deliver a multi-model ensemble result that would be statistically indistinguishable, as far as both global and regional outcomes are concerned.

Line 231: this should be qualified as “CO2 concentrations”.

Thank you, done.

Lines 236-238: this is similar to when, for example, aerosol emissions and cloud-aerosol interactions were included. What is the approach suggested to help with the challenge in the interpretation of the results? Who will be helping?

We are asking the E-driven participants to run *at least* one of the scenarios also in C-driven mode (we recommend the Medium scenario). We hope that this will help apportion the uncertainty to the different sources. We are aware of the possibility that the results would be dependent on the forcing levels, hence the “at least”, and we specify two more scenarios in concentration mode in Tier 2, but given the computational burden imposed on modelling centers we cannot require more.

Lines 252-254: shouldn’t there be a definition of which scenario? Is the implication that the concentration-emission comparison is scenario independent? That seems hardly justifiable.

Since up to this point we have not described the scenarios, we are not mentioning which, but we will specify it later. We are also hoping the modeling centers will run more than one scenario in both modes, but computational constraints prevent us from requesting that. We agree that the comparison may be scenario dependent. Within the paper, ScenarioMIP is suggesting scientific directions of future research to the community, and this is one of those.

Lines 260-262: this would be a step backwards for most climate models since cloud-aerosol emissions are now a critical process that needs to be represented. There are multiple implementations of reduced chemistry relevant to aerosol only.

We have edited this paragraph to clarify our request to modelers, which is to run CO_2 as emissions-driven, other well-mixed GHGs as concentration-driven, and to provide aerosol and ozone precursor emissions so that models with atmospheric chemistry can run in emissions-driven mode. At the same time, we will provide ozone concentration and optical aerosol properties for those models that need to run these species in concentration-driven mode.

Line 264: what is proposed here in terms of AerChemMIP? To use the output of ESMs run with full chemistry? Very confusing sentence.

We agree the sentence is confusing. Since the topic of future research directions is already covered in the concluding section, we have deleted the sentence here.

Lines 268-269: observations of what? CO2?

We're refining this point to be less specific on what we demand of modeling centers with regard to their historical CO_2 concentration outcomes.

We replaced those lines with the following text:

"Running historical simulations in an emission-driven configuration presents a challenge, as a model's simulated CO_2 concentration can drift from observations. However, this approach is a deliberate and core feature of the CMIP7 experimental design. Rather than take measures to artificially correct this bias, we encourage modeling groups to quantify CO_2 concentration evolution as a key diagnostic of model performance and carbon cycle sensitivity. To facilitate this analysis, the ScenarioMIP protocol includes a parallel concentration-driven simulations specifically to allow for a clean diagnosis of the impact of these emergent carbon cycle feedbacks on climate outcomes."

Line 272: why still stop at 2100? IAMs have been used over the last couple of decades and have always had 2100 as their end point. One would therefore conclude that IAMs could be used for 100-year projections. Why not here?

IAM simulations depend on driving datasets (e.g., future time series of population and GDP at country-level resolution) that have not been produced yet beyond 2100. In this

phase of CMIP, given the tighter than usual timeline, it has not been possible to work on that aspect and explore IAM's performance for a longer timeline.

Line 278: emissions are not observed, only estimated.

Reworded, thanks for pointing that out.

Line 285: why that request of 2150? Is there the expectation that something interesting will happen between 2100 and 2150? This needs to be justified

The request to run simulations out to 2150 is to allow for a longer time horizon for adaptation planning and for the assessment of overshoots with all ScenarioMIP participants' contribution. Scenario experiments were first conducted in the 1990s, with a 2100 time horizon. A similar projection length from present day requires advancing the time horizon of the simulation, getting closer to 2150. While early plans for ScenarioMIP in CMIP7 were for IAMs to be run out to this date, technical barriers prevented this from happening as we mentioned above. It was thus decided that, as it remains valuable to have a large number of ESMs extend the standard simulations out to 2150, the scenario extensions' forcings could be used for this segment. This will enable the assessment of climate commitments, and greater focus on long term impacts and reversibility of an overshoot, for example.

Line 292: why is there a need for a new high emission scenario? Can't we just re-use SSP5-85 (or SSP3-70) with updated harmonization? Looking at Fig 1a, the upward bend in CO2 emissions for this scenario seems hard to justify, unlike the M and ML scenarios

As we know, the 8.5 (and even 7.0) scenario have been criticized now for a long time. In addition, many conditions relevant to projections have changed, including most notably costs and rates of capacity addition of renewables, base-year data, anticipated trends in technology adoption and costs, GDP estimates, etc. Also, the CMIP6 SSP3-7.0 had higher loading of aerosols that were seen as "masking" warming and affecting the hydrological cycle, thus creating challenges in interpreting climate outcomes. The impact community asked for *as high a scenario as plausible* in terms of warming, not just CO_2 emissions, thus our new high scenario is not going to have high aerosol loading.

Line 315: it is never “exact”, unfortunately

Agreed, we reworded as “actual.”

Line 335: what will be the process to decide which emissions to use? Will ESMs be allowed to use whichever IAM output they want?

A single marker scenario will be chosen as in the past, and the choice and its rationale will be documented in papers that the IAM teams plan to publish in 2026, detailing the makeup of the various trajectories.

Lines 356-360: the recent IMO regulations of sulfur emissions have been shown to exert a very significant radiative forcing. See https://doi.org/10.5194/acp-24-13361-2024 for example

We are talking specifically about regionally detailed outcomes across ESMs, but we have added the point about global ERF and the citation, thank you.

Line 367: who will be in charge of creating this set of emissions? SSP3-70 and SSP3-LowNTCF played that role in CMIP6

At the moment, there are no specific plans to run alternative scenarios in AerChemMIP. In the text, we now have left this open - if future requests would arrive. We have some variations in ScenarioMIP itself, as we have now more clearly indicated.

Lines 383-384: why only involve the IAM community?

That’s a fair question, but the reality is that these would be highly technical and IAM-centered evaluations of the consistency and assumptions behind the various pathways, characteristics that would be hard for a different community to assess.

Lines 424-442: this discussion does not seem to add anything useful to this paper. I suggest deleting

We respectfully disagree, as the role of emulators is becoming more and more important, and, through interactions with the CMIP chairs, this was identified as a topic to specifically address.

Table 2: does this table suggest that ScenarioMIP will be in charge of generating concentrations of all gases listed, aerosol optical depth (which by the way is not a very useful diagnostic to drive a climate model) and ozone?

We have edited the table to make this clearer. Where not otherwise noted, the emissions and concentrations listed will be provided by ScenarioMIP. Most gases' concentrations will be obtained using a Simple Climate Model. Aerosol optical depth and ozone concentrations will be provided by an atmospheric chemistry model that will be using IAM emissions as input.

Lines 461-463: But the DECK only looks at the response to CO2 forcing and associated feedbacks. Other MIPs should be in charge of other elements of comparison.

We stand by the decision not to be concerned with this aspect explicitly. In the past, experiments or even additional MIPs, like the reviewer suggests, were specifically designed to explore consistency of past and present phases, but the uptake of the question and the experiments by the community did not seem significant enough to warrant taking this as a constraint for the current design, and as a role for ScenarioMIP.

Lines 517-519: what kind of extreme events would lead to high emission future?

We agree this was opaque and we have deleted these lines.

Line 532: what is meant by "strong" here?

We reworded as "strongly argued and well-founded storyline"

Lines 664-667: why is CDR measures as part of this discussion? If ESM produced temperature signals can't tell the difference between the scenarios, why would those

scenarios be run? This seems more like a topic of interest to the IAM community, but not ScenarioMIP. Maybe part of another MIP?

We have designed the scenario VL to be different enough from LN. Differences can come from CO2 emission pathways, non-CO2 gases, land use as well as from things related to CDR (both the biophysical consequences of land use and also the different representations of CDR in ESMs). However, as in the evaluation using small climate models we only look at the first two factors and the current formulation is confusion, we have now decided to simplify the text and only mention the two first factors..

Line 734: how will forcings be harmonized? Or is it meant to be emissions and LULCC that will be harmonized?

Yes, correct and corrected, thanks.

Line 743: replace “small” with “simplified”

Thank you, this was an oversight of course.

Line 834: I don’t believe CDR is part of GeoMIP

Thank you for pointing this error out, we have indeed learned from the MIP site: “*We are specifically not studying Carbon Dioxide Removal (CDR) methods, wherein carbon dioxide is removed from the atmosphere; potential methods include enhanced mineral weathering, ocean iron fertilization, or direct air capture. Model intercomparisons of CDR are being investigated by the Carbon Dioxide Removal Model Intercomparison Project (CDR-MIP)”* and we corrected our text accordingly.

Line 856: how would ESMs use BECCS resulting concentrations? ESMs don’t have a CO2 tracer just for BECCS.

We have clarified the sentences. The BECCS net result in terms of emissions will be part of the information on CO_2 emissions provided by the IAMs, so ESMs won’t process them separately but will respond to their effect as part of the overall outcome in total

emissions. The mentioning of concentrations in this sentence referred to ESM models that cannot calculate concentration. These models will directly use the resulting CO2 concentrations calculated on the basis of IAM output (including emissions and CDR). However, directly coupling the word concentrations to CDR created a wrong impression.

Lines 867-869: doesn't enhanced weathering depend on plant productivity (https://www.nature.com/articles/s41586-024-08429-2)?

Although enhanced rock weathering is performed on agricultural fields in this example, it does not significantly depend on plant productivity. It could influence plant productivity, though, via minerals and/or impacts on ozone. However, because for this section of the text the impact on plant productivity is not relevant, we have removed this part of the sentence.

Lines 874: do IAMs have a representation of wildfires? ESMs will definitely have that in the emission-driven runs. This could lead to a significant mismatch

There are several reasons why there might be a mismatch, including the different growth rates in IAMs and ESMs, different saturation rates, the impact of climate change and also wildfires. In the experiments we hope to learn from these differences. We have now mentioned wildfires as a possible source of difference. In IAMs, wildfires might be part of the calibration of the forest growth rates (but indeed typically do not include a process representation).

Lines 949: actually, this paper has proposed a limited set of scenarios (section 3), which, once modeled by IAMs, will lead to emissions, concentrations and LULLC. Not the way it is stated

We have edited the text to correct this statement.

Lines 971-976: again, this ignores the fact that many ESMs have an explicit representation of aerosols. Furthermore, why would "a single atmospheric chemistry

model" be the right approach? Which one would ScenarioMIP pick? It seems that a tighter collaboration with AerChemMIP is required

Please see our response to the comment above on lines 260-262, which addresses the recommendation for emissions- vs concentration-driven runs. We edited the paragraph in section 2.2 to clarify our request to modelers. Since that request involves providing emissions of aerosol and ozone precursors so that ESMs can run these species in emissions-driven mode, we have deleted the text referred to here, which is no longer relevant.

Lines 977-982: are the authors arguing for the use of RCP8.5 as it was designed for CMIP5, or for SSP5-85 for CMIP6? Are those going to be harmonized to the same 2025 emissions/concentration data? What about their extension to 2500? Is this part of CMIP7? Very confusing

We agree this point is confusing and have deleted it.

Nathan Gillett (CC2):

Overall I found the article very interesting, and I think the overall experimental design of ScenarioMIP for CMIP7 is good. I think the proposal to focus on emissions-driven simulations is a good one. I applaud the co-chairs for the comprehensive consultation process and their diverse author team. My comments focus on particular aspects of experimental design and narrative.

The authors include a box (Box 1) on characterizing scenario likelihood. The box includes reference to likely scenarios, which have a relatively high probability of occurring. However, the authors state on lines 539-540 that the medium emissions scenario should not be considered a most likely scenario. The authors do not cite or discuss the substantial recent literature on probabilistic scenarios (e.g. Sarofim et al. (2024, https://www.nature.com/articles/s41467-024-52437-9), Moore et al. (2022, https://www.nature.com/articles/s41586-022-04423-8,), in which probabilities are assigned to future emissions, including papers which explicitly aim to characterize the likelihood of previous-generation ScenarioMIP scenarios e.g. Huard et al. (2022, https://agupubs.onlinelibrary.wiley.com/doi/full/10.1029/2022EF002715). The latter

study calculates likelihoods for each of the SSP-based scenarios used in the previous generation of ScenarioMIP, based on the results of other probabilistic scenario studies, and for example finds that intermediate scenarios, like the medium scenario, are more likely than some other scenarios. Note that such probabilistic scenarios offer substantial potential benefits for use in adaptation planning, since they can be used to make probabilistic projections which are not conditional on particular emissions scenarios. I am not suggesting that ScenarioMIP actually includes such ensembles of probabilistic scenarios, but I think it would be good to cite and discuss this literature to situate the MIP relative to this recent research. For example the authors could cite and discuss this literature and the potential benefits of such probabilistic scenarios for informing adaptation, and say that while they encourage individual modelling groups to run such scenarios, computational constraints preclude running ensembles of emissions across multiple models in ScenarioMIP – and the focus here is on sampling over model uncertainties for a limited range of plausible scenarios.

We agree that it is worthwhile to refer to the probabilistic projections literature and make clear how the ScenarioMIP approach to uncertainty relates to it. We have added a paragraph to the text in section 2.2, subsection on "a wide and plausible range," that refers to recent literature ascribing probabilities to emissions projections (note that Sarofim et al. and Huard et al. do not develop probabilistic emissions projections, instead using projections developed in other studies, so we do not include these two references here). We note that we don't take such an approach in the ScenarioMIP design, given the lack of wide agreement on methods and the necessity of subjective judgments. We believe this topic is better left as an open research question. However, we do now indicate that given the nature of the experimental design in which the highest and lowest emissions scenarios are intended to push the boundaries of plausibility, we anticipate that these bounding scenarios would be less likely than the others.

While I agree with the point, raised in Box 2, that different assumptions about inequality between regions, reflected in different regional patterns of emissions, will not strongly affect the climate outcome, other aspects of physical climate change might be more sensitive to equity and justice assumptions. For example, integrated assessment modellers typically choose a cost optimization approach to reaching a certain level of radiative forcing in 2100, based on a discount rate which embodies intergenerational justice assumptions, and assumptions about the availability of CDR

approaches in the future, many of which require extensive land (for example for afforestation, or growing bioenergy crops for use in BECCS), which may in turn impact food security or Indigenous land rights (e.g. Rubiano Rivadeneira and Carton, 2022 https://www.sciencedirect.com/science/article/pii/S2214629622002845). These scenarios typically offset relatively higher fossil fuel emissions in the near term with relatively larger CDR towards the end of the century. Different assumptions about intergenerational equity (i.e. discount rate), or availability of land for CDR, might change these scenarios in a way which would affect global climate evolution. Moreover the use of extensive CDR in the SSP-based scenarios consistent with Paris targets used in the previous iteration of ScenarioMIP is itself often used as a justification for the development of CDR – even though as just described other assumptions in the IAMs might have resulted in less CDR. Similarly, some papers argue that the strong use of BECCS in IAM scenarios is a product of the IAM structure and assumptions, and directly led to the stronger consideration of BECCS by the policy community (e.g. Creutzig et al., 2021 https://onlinelibrary.wiley.com/doi/10.1111/gcbb.12798). Overall, the current draft slightly gives the impression that the authors are trying to sidestep equity and justice considerations by arguing that the physical climate is not very sensitive to these assumptions. But even the selection of certain scenarios for inclusion in ScenarioMIP, with certain ranges of emissions, CDR, BECCS etc can send a message to the policy community, even if this isn't the authors' intention. So I suggest that the authors include a stronger caveat in this box saying that the scenarios selected for inclusion in ScenarioMIP reflect certain assumptions in the underlying IAMs, particularly with regard to equity and justice issues, and should not be seen as a full range of possible futures, particularly with regard to technologies such as CDR and BECCS.

We agree that there are more assumptions in IAMs (including discounting and the availability of CDR options) that have equity implications, so we have added text to Box 2 as suggested by the reviewer. There is also a clear indication to further research on equity-related aspects in future scenarios.

I was interested to see that the VLLO scenario may include ocean-based CDR strategies (ln 643-644), by which the authors presumably are thinking of geoengineering techniques such as ocean fertilisation or ocean alkalinisation (given that the potential carbon uptake of coastal measures is limited). Such measures stimulate uptake of CO2 by the ocean by altering ocean biogeochemistry on the large scale, and in ways which might interact with climate change induced changes. Also such measures could have

implications for ocean ecosystems and ocean chemical/physical properties, including changes which could be simulated by ESMs. However, according to lines 247-250 “only afforestation and reforestation will be based on endogenous representation of land - based mitigation solutions in ESMs. For all other CDR options we will include their emission impact within the IAM emission output”. This seems like a potential miss from a science perspective to me. Also, omitting direct biogeochemical effects of ocean-based CDR in the simulations would mean that variables such as ocean pH or ocean carbon uptake, which are regularly examined and reported on including in IPCC reports, would not be correct in these scenario simulations, because they would be missing important biogeochemical processes (ocean alkalinization directly alters ocean pH, and both alkalinization and fertilization aim to directly increase ocean carbon uptake – whereas specifying the effects of these measures by simply reducing net global CO2 emissions would reduce ocean carbon uptake). I suggest either not including ocean-based CDR in this scenario, or if it is included, then either it should be modelled explicitly in ESMs, or if not then the authors need to include discussion of the limitations of not doing so. Separately, I think the inclusion of such measures in the VLLO scenario would need to be treated carefully to avoid the interpretation that only scenarios with ocean alkalinization/fertilization were able to keep global temperatures below 1.5°C by the end of the century.

Many thanks for this comment. We agree that the VLLO and VLHO scenarios (now renamed as VL and LN, Very Low and Low-to-Negative, respectively) in particular would be strong candidates for a CDR component in the narrative, and it is a high scientific priority to simulate CDR from both land and ocean strategies in order to model both their process-resolved efficacy as carbon removal and their potential impacts on other elements of human and natural systems.

However, the implementation of these CDR strategies raises technical challenges in terms of scenario design (many IAMs do not currently fully represent ocean CDR as part of their mitigation portfolio), data pipeline (the data request to unambiguously and consistently define removal fluxes across different and diverse IAMs), downscaling (the mapping of coarse regional actions in the resolved IAM to spatially meaningful distributions for use in the ESM), ESM CDR process implementation, and sequestration and the reporting of sequestration fluxes within the ESM/CMIP data request.

None of these technical challenges is impossible, but the tight timescale of the ScenarioMIP process prevents us from being able to realize this dimension within the core ScenarioMIP set. We hope, however, to work with CDRMIP to produce resolved CDR scenario experiments within the CMIP7 cycle, allowing for a small amount of time to implement the technical pipelines required to meaningfully represent both land and ocean CDR in the ESMs.

The text on lines 709-732 motivates extending simulations beyond 2100 based on understanding long-term dynamics of the earth system and exploring reversibility. Based on these considerations idealized extensions of the scenarios are proposed to 2500. These idealized scenarios include unrealistic features such as abrupt changes in the rate of changes of emissions, and extremely high levels of cumulative negative emissions (around 6000 GtCO2 in H-ext-OS) (Figure 2). But increasingly climate projections beyond 2100 are needed for climate adaptation (e.g. Lyon et al., 2024; https://pubmed.ncbi.nlm.nih.gov/34558764/ ; Park et al., 2025 https://www.science.org/doi/10.1126/sciadv.adn8819; Easterling et al., 2024 https://www.nature.com/articles/s41558-024-02085-0). And 2100 will only be about 70 years away by the time much literature based on the ScenarioMIP simulations is published. Literature containing plausible emissions scenarios beyond 2100 does exist - for example Sarofim et al. (2024) https://www.nature.com/articles/s41467-024-52437-9 contains a probabilistic scenario to 2300 based on expert elicitation. Meinshausen et al. (2020) contained projections to 2500 for the SSP scenarios https://gmd.copernicus.org/articles/13/3571/2020/, based on assumptions of emissions evolution for each sector. Even the previous generation of ScenarioMIP used more plausible scenario extensions to 2300 than those proposed here (O'Neill et al., 2016; https://gmd.copernicus.org/articles/9/3461/2016/gmd-9-3461-2016.pdf). Even though uncertainties are no doubt larger for scenarios extending beyond 2100, I think it would be a disservice to the impacts and adaptation communities to choose scenario extensions beyond 2100 which are obviously implausible – for example imagine trying to defend coastal defense or mine reclamation plans post-2100 based on one of the scenarios in Figure 2. I encourage the authors to reconsider the extensions beyond 2100 and update with scenarios which are plausible and consistent with the assumptions underlying the scenarios up to 2100. And/or, if the authors retain some idealized scenarios which are not intended to be plausible, then I suggest that they

clearly flag these so that they can be separated from those scenarios which are intended to be plausible post-2100. (I notice that Jean-Francois Lamarque raises a related issue in his comments).

Thanks indeed for your comments and for your well reasoned advocacy on behalf of the impacts and adaptation communities. We take your point that providing plausible, realistic scenarios beyond 2100 is a critical goal, and we agree that the initial design could be improved.

Firstly, though we agree that physical-societal evolution significantly beyond the 2100 timeframe should be a fundamental concern for society, the IAM community does not consider current integrated assessment models fit for purpose for multi-century scale applications. Economic discounting assumptions which lie at the core of IAMs, combined with technological assumptions which are fundamentally grounded in present day systems, become less appropriate with longer lead times, and for this reason, the majority of IAMs do not attempt to simulate a multi-century timescale. As such, meaningful models of climate-human-ecosystem evolution beyond 2100 would be a fascinating tool, but we would not consider that this can be achieved by simply extending the time horizon of existing IAMs for centuries.

On the more pragmatic question of whether the time horizon can be extended somewhat to account for the shortening timeframe to 2100 specifically, we fully agree that this would be desirable, and in the early stages of the ScenarioMIP planning process there was ambition to explore a 2150 end date for IAM simulations. However, achieving this in practice requires assembling a large library of input data/drivers which define time-dependent boundary conditions; for example, defining scenario-dependent technological cost evolution, population and GDP trajectories with country-level resolution. Unfortunately, the tight timescale of the ScenarioMIP process precluded this plan from being finalised in time for ScenarioMIP CMIP7, but it remains an ambition for future scenario development and climate model intercomparisons.

Given this, for CMIP7, the role of the post-2100 extensions is largely to provide a framing for emissions trajectory extrapolations which are continuous with the 21st century scenarios and span a range of futures which are both plausible and of scientific interest to diverse communities. As you note, the impacts and adaptation communities are key examples, but the extensions are also of interest for fundamental understanding of carbon-climate dynamics on a range of timescales which can help to

inform process understanding and guidance on the limitations of policy tools such as carbon budgeting based on the cumulative emissions accounting.

Based on your feedback, we have made two major changes which can allow us to increase the plausibility of our scenario extensions while staying within the necessarily idealised framework in which we're constrained. First, we have removed the most extreme overshoot scenario (Hext-OS), from the protocol and replaced it with HLext, which represents a smaller and more plausible transition from a High scenario to one aligned with long-term Paris Agreement goals.

Second, we have added a new section to the paper (4.6) that comprehensively addresses the issue of "unrealistic features such as abrupt changes." Across all extensions, we have implemented methodologies to ensure smooth transitions from the 21st-century pathways and updated the figures with these new results. This includes ensuring continuous derivatives for both CO_2 and non-CO_2 emissions. We are actively attempting to implement land use assumptions which will not result in unrealistic step changes (an iterative process with the LUH2 team). We hope that these revisions will result in a set of extensions that better serve the communities you highlight.

Detailed comments:
Ln 268-269: How are modelling groups expected to "keep deviations from observations in the historical period to a minimum"? Does this refer to CO2 concentration alone, or CO2 concentration and temperature? As written it sounds like the authors might be suggesting nudging temperature and CO2 towards observations or similar, but I don't think this is part of the historical experimental design. Probably the authors mean through model tuning – if so, state this explicitly. Also, the use of historical climate change to tune climate models is controversial, with many groups aiming to only tune the present day climate, and future warming being an emergent property of the model arising from the physics of the system, not something which is set by tuning (e.g. Hourdin et al., 2017 https://journals.ametsoc.org/view/journals/bams/98/3/bams-d-15-00135.1.pdf). Are the authors endorsing tuning models to reproduce historical warming here?

Thanks for this point. We agree that this text presented an unrealistic request for model development teams.

We have replaced the text with what we hope is a more nuanced description of the challenge introduced:

“Running historical simulations in an emission-driven configuration presents a challenge, as a model's simulated CO_2 concentration can drift from observations. However, this approach is a deliberate and core feature of the CMIP7 experimental design. Rather than take measures to artificially correct this bias, we encourage modeling groups to quantify CO2 concentration evolution as a key diagnostic of model performance and carbon cycle sensitivity. To facilitate this analysis, the ScenarioMIP protocol includes at least one parallel concentration-driven simulations specifically to allow for a clean diagnosis of the impact of these emergent carbon cycle feedbacks on climate outcomes.”

Ln 269-270: The authors argue that it is expected that analysis of future projections will focus on deviations compared to the start year of the simulations. This was not generally the case for analysis of future projections in the IPCC AR6 WGI report, because of the interest in climate change relative to preindustrial, as referenced in the Paris Agreement. For example, Figures SPM 4, 5, 7, 8 and 10 in the summary for policymakers of the AR6 WGI report all use quasi-preindustrial baselines for projections.

We have deleted that statement in the process of rewording the entire paragraph where it belonged, as just described.

Ln 299: The text refers to the high emissions scenario as reflecting “slow development of mitigation technologies and diffusion”. My understanding is that high emissions scenarios such as SSP3-7.0 or SSP5-8.5 reflect no new mitigation actions at all, and as just stated on line 293, a reversal of trends towards mitigation. I recommend repeating the language concerning a reversal of current trends towards mitigation here.

We agree this text was confusing and we have removed it, adding the reference to reversal of mitigation policies as the reviewer suggests.

Ln 378-388: The timeline for provision of forcings from the IAMs seems a bit vague here, and the text makes it sound hypothetical and far in the future. But on lines 474-

475 the authors say that the IAMs are running over the period September 2024 – June/August 2025, in order to meet the needs of IPCC AR7. So the IAM runs are presumably well underway. Can the authors add more here on the progress on the IAM runs so far? Are any outputs of early IAM runs available for inclusion in this paper? I find it a bit unsatisfactory not to have actual IAM outputs included in this paper – the equivalent description paper for CMIP6 (O'Neill et al. 2016) did include actual forcing timeseries from IAMs.

This issue has become even more important by now, as the IAM scenarios are now close to being final. However, as they are not final yet, and will be published in papers of the IAM community (with a much more detailed description), we have not included them in this paper. We have, however, revised significant parts of the paper based on the current status of the process. We have therefore produced new versions of the figures showing illustrative emissions pathways and their results concentrations, forcing, and temperature change, based on preliminary IAM and SCM results. We have also edited the paper throughout to change the perspective so that it is from the current scenario status. This has led to edits most prominently in sections 1 (introduction), 2.3 (scenarios), and 2.4 (design issues related to IAM runs).

Ln 512: A 'competition scenario case' is mentioned here, but this term hasn't been introduced, and there is no description of what this means.

The text has been changed significantly, and the term is no longer used. .

Ln 517-518: Are the extreme events referred to here climate/weather extreme events? If so, this statement that extreme events could lead to high emissions seems to contradict what the authors say on lines 210-211, and again immediately following this on lines 521-524 – namely that all the scenarios used assume no climate impacts.

We agree the text was confusing and have deleted it.

Ln 589-591: This text references uncertainty in the carbon cycle response as a source of uncertainty in whether or not the global mean temperature can be returned to 1.5°C. But uncertainty in the physical climate response is also very important here, given for

example the still relatively large uncertainty in equilibrium climate sensitivity. I would guess that physical climate uncertainty might be even more important than carbon cycle uncertainty for determining whether or not 1.5°C at the end of the century can be achieved.

Fair point, we have reworded as "(although it should be noted that the actual behavior of the ESMs is at the moment uncertain, due to both uncertainty in the physical climate response and the added uncertainty in the carbon cycle response, when models are run in emission driven mode). "

Ln 665-667: The text here refers to additional impacts from CDR measures. But according to lines 247-250 only afforestation and reforestation will be represented endogenously. Do the authors just mean additional impacts of afforestation and reforestation here? Or if not, how will these impacts be represented in ScenarioMIP? Clarify.

The scenarios include afforestation, but also other CDR measures. In the ESMs afforestation will be represented explicitly and thus indeed, the other implications of afforestation were referred to in this sentence. However, for other CDR measures the response in ESMs to net negative emissions may also lead to scenario differences. The wording of the sentences has been rephrased to make it clearer.

Ln 682: The text here says "the mechanisms and extent of CDR deployment will have ESM-specific efficacies", but lines 247-250 say "only afforestation and reforestation will be based on endogenous representation of land -based mitigation solutions in ESMs. For all other CDR options we will include their emission impact within the IAM emission output". If these statements are both true, then the first statement must surely refer only to afforestation and reforestation, since all other CDR options are apparently just represented by reducing anthropogenic CO2 emissions. Please clarify.

We have made the text more precise. Indeed, afforestation and reforestation are represented process-wise in ESMs.

Ln 748-751: Is H-ext really plausible as stated here? There is now extensive literature on the low likelihood of SSP5-8.5 (see Chen et al. (2021), https://www.ipcc.ch/report/ar6/wg1/chapter/chapter-1/, Hausfather (2025)

https://journals.sagepub.com/doi/10.1177/29768659241304854, Sarofim et al. (2024) https://www.nature.com/articles/s41467-024-52437-9 among many other sources). This scenario has a similar level of radiative forcing in 2300 as SSP5-8.5 (Meinshausen et al., 2020).

Several impact research areas are interested in the H-ext extension to explore the long term Earth System response to emissions above current policy estimates. Please also note that our H scenario is significantly lower than SSP5-8.5 in emissions, and therefore expected forcings and temperature outcomes.

Ln 748-749: I would encourage the authors to consider making the M-ext simulation a high priority simulation. Recent literature suggests that emissions of around the medium level are most likely (e.g. Sarofim et al. (2024) (https://www.nature.com/articles/s41467-024-52437-9), Moore et al. (2022) https://www.nature.com/articles/s41586-022-04423-8, Huard et al. (2022) (https://agupubs.onlinelibrary.wiley.com/doi/full/10.1029/2022EF002715), Hausfather (2025) https://journals.sagepub.com/doi/10.1177/29768659241304854). And as discussed above, many adaptation planners will require scenarios post-2100. Relying only on H-ext and VLLO-ext (the two planned high priority extensions) means relying on two low likelihood/less plausible scenarios.

In our interaction with the impact research community it was found that the two bracketing extensions remain the most desirable. Also, please note that the experimental design requires all scenarios, and therefore the M scenario too, to be run out to 2150 using the first 50 years of the extensions forcings. Therefore, we hope to balance service to the impact community and computational demands on modeling centers by exploring the extremes of the range, but also providing more than 100 years' worth of projections according to the M scenario among all others.

Ln 977-982: The authors might also note that consideration of concentration-driven SSP5-8.5 simulations was in the past motivated by the possibility that stronger carbon-climate feedbacks than those included in MAGIC might give rise to SSP5-8.5 CO2 concentrations even if emissions were lower (e.g. Chen et al., 2021, IPCC AR6 WGI). This time around ScenarioMIP is recommending emissions-driven simulations, so even though the High scenario emissions are below those of SSP5-8.5, the CO2

concentrations could be at similar levels in models with a high airborne fraction/strong carbon-climate feedbacks.

Thank you, this was a great point, but no longer relevant as the text has been deleted.

Ln 1028-1033: Another area where emulators may be useful is for simulating future climate change under probabilistic ensembles of scenarios. The authors could mention this here.

Great point, now added to the text.

Florian H (CC3):

The current ScenarioMIP proposal introduces an inconsistency in how CO_2 emissions from land-based Carbon Dioxide Removal (CDR) methods are handled. Specifically, while ESMs take IAM-derived CO_2 emissions for Bioenergy with Carbon Capture and Storage (BECCS) as given, they recalculate CO_2 emissions related to afforestation and reforestation instead of using IAM-derived estimates. This approach creates an internal inconsistency in the treatment of land-based CDR options.

Explanation of the Inconsistency:

IAMs generate internally consistent land-use and emissions pathways, balancing land competition between food production, bioenergy crops, re/afforestation, and conservation while also projecting associated CO_2 fluxes. However, in the proposed framework:

- BECCS-related CO_2 emissions and removals are directly taken from IAMs and used in ESMs without modification.
- Afforestation and reforestation CO_2 emissions/removals are recalculated dynamically in ESMs, considering soil carbon dynamics, nutrient limitations, and climate-vegetation interactions.

This discrepancy creates an inconsistent approach to land-based CDR—while BECCS emissions are assumed to follow IAM projections exactly, afforestation and reforestation are subject to biophysical recalculations that may significantly alter the land-use CO_2 emissions trajectory. As a result, the total carbon budget and net-negative emissions estimates may diverge from what IAMs initially projected.

Additionally, many IAMs already include a process-based representation of forest growth and carbon sequestration(i.e., the biogeochemical effects of afforestation and reforestation), meaning that they dynamically simulate carbon uptake by vegetation. However, IAMs generally lack a representation of biophysical effects such as changes in albedo, evapotranspiration, and surface energy fluxes, which can significantly impact regional and global climate. These biophysical effects are the strength of ESMs and should be incorporated in a way that complements, rather than overrides, IAM-based carbon sequestration estimates.

Suggested Improvement:

To ensure internal consistency in the treatment of land-based CDR, while still leveraging the complementary strengths of IAMs and ESMs, I suggest the following refinements to the proposal:

1. Use IAMs as the primary source for all land-use CO_2 emissions projections, including both BECCS and afforestation/reforestation, ensuring coherence in how CDR is represented.
2. Use ESMs to simulate the biophysical feedbacks of land-use changes (e.g., climate-vegetation interactions, albedo effects, surface temperature changes, and hydrological responses) without recalculating CO_2 emissions from afforestation and reforestation.

This approach would ensure a more coherent and transparent integration of land-based CDR into ScenarioMIP, preventing discrepancies in carbon budget estimates while maintaining the policy relevance of IAM scenarios and the biophysical accuracy of ESM outputs. I strongly recommend incorporating this refinement into the ScenarioMIP proposal.

Beyond ScenarioMIP, an iterative feedback mechanism, where ESM-derived insights on biophysical feedback effects are used to refine IAM land-use and emissions projections could be established in the future.
Citation: https://doi.org/10.5194/egusphere-2024-3765-CC3

Many thanks for this suggestion. We fully take your point that the current pipeline will introduce an inconsistency between the resolved carbon fluxes associated with afforestation and other land management activities as they are resolved within the Integrated Assessment Models and the Earth System Models.

However, we consider this to be a critical part of the emissions-driven experimental design for ScenarioMIP CMIP7. There is - as you note - a disconnect between the process representation in the land surface models in IAMs and ESMs in both directions - the IAMs exclude climate-biosphere interactions, while the ESMs often exclude large refinement of energy crops and effects of technological investment in yields.

As such - we see a role of the ESMs in this system as being a source of objective uncertainty in the land use emissions outcomes from land use decisions outlined in the IAM. Just as the Earth System Model will simulate a different global mean temperature outcome to the original IAM, we expect that consequences of an afforestation strategy - for example - will have a range of consequences for land carbon uptake (or loss) in the ScenarioMIP ESM ensemble.

ESMs will already be computing emissions associated with land-use change in the historical simulations (e.e.g expansion of cropland, wood harvesting, etc). It makes sense to continue this in the future scenarios, now with the addition of land-use change associated with land-based CDR such as afforestation. In fact it would be very difficult to separate emissions reductions from afforestation in an ESM from other land-use change emissions. Although the ESMs will calculate emissions associated with land-use change from expansion of bioenergy crops, the emissions reductions from the associated CCS cannot easily be computed by most ESMs at this point, so it makes most sense to have this component of BECCS emissions provided by IAMs.

Massimo Tavoni (CC4):

The current framework nicely explores a wide range of emissions and associated temperatures. However, political turmoil has intensified since it was conceived a year ago, with significant backlash against mitigation measures emerging. As a result, I think it would be great to expand the scenario by reflecting an additional narrative that explores higher ranges of climate overshoot.

Currently, temperature peak and decline is explored in the VL scenarios and possibly also in ML. Both are important, but they currently cover a limited range of future possibilities.

The VL scenarios are in the context of Paris-compliant policies stabilizing temperatures at 1.5°C. Though clearly important, they do not reflect a scenario with very limited mitigation action in the next years to decades. The ML scenario is a very important one, but the extent of the overshoot is limited and mainly occurs during the next century. Though extensions beyond 2100 explore various degrees of overshoot, the timescales are so long that they will not be policy-relevant. Also the uncertainties on socio-economic dynamics over such timeframes are so significant that they make it hard to interpret results.

Based on these considerations, adding a seventh scenario with sustained emissions followed by a late awakening of the gravity of climate change followed by rapid action to bring emissions towards zero would be important. Such a HL scenario could have the following characteristics:

> Follow H till after mid-century (e.g. 2060/70), then try to hit net zero CO2 by 2100/2120
> This will result in a significant overshoot of 1.5°C warming followed by very rapid drop in temperature (e.g. 0.5°C in a few decades)
> Such a scenario will provide an interesting, disruptive scenario narrative illuminating the risks and consequences of a sustained political backlash on mitigation
> The scenario will be relevant for earth system science because it will highlight the carbon cycle and climate repercussions of rapid CO2 drawdown and the climatic risks of overshoot
> It will also be relevant for the climate impact community to help quantify the physical risks of high-temperature overshoot
> And it will be relevant for the socio-economic communities to understand the social and economic consequences of overshoot (including for key issues such as adaptation needs and loss&damages)

Citation: https://doi.org/10.5194/egusphere-2024-3765-CC4

Thanks for this suggestion. We agree that a High-Low scenario would be useful for exploring the consequences of late mitigation following a high emissions path, leading to a high temperature overshoot outcome. We have added such a scenario to the ScenarioMIP design. We have adopted the suggested design feature of emissions declining to net-zero CO_2 by 2100, but note that this would not lead to temperature

overshoot this century, but only after 2100 (when the extension of this scenario includes net negative CO_2 emissions).

Roberto Schaeffer (CC5): +2!

1. It would also be relevant for the climate mitigation community to help it quantify the transition risks of rapid increase and then decrease in emissions, and
2. It would also be relevant for the climate mitigation community to better explore some equity implications of such a scenario.

Thanks, the final design has taken up this suggestion and we have now an HL scenario

Colin Jones (RC2):

This paper outlines a proposal for a set of emission and land-use scenarios to be developed by the international Integrated Assessment Modelling (IAM) community, which will act as input to Earth system models (ESMs) in CMIP7. The emission scenarios (and therefore ScenarioMIP) are a central component of CMIP7 and one of the primary routes that new ESMs are used, sampling a common set of future emission and land use scenarios, to generate an internationally coordinated set of future climate projections. The resulting projections support numerous downstream activities across science, policy, and real-world decision-making. ScenarioMIP is therefore a key foundational activity for climate research, for climate policy, and for climate change decision making (mitigation, adaptation, and increasingly modification).

The ScenarioMIP project, with this paper as one important outcome, is the result of a long, detailed, open, globally inclusive, and highly responsive consultation. For this the leaders of ScenarioMIP deserve enormous credit as it has required significant time, effort, patience, and expertise. The degree of inclusivity, and the willingness to incorporate views from numerous disciplines and from regions of the world, will make the resulting emission scenarios and the projections based on them significantly stronger and of great utility across the world. The paper therefore clearly needs to be published and as it stands now is very close to publication. I have a few concerns and suggestions that I outline below. I hope these contribute to making the paper stronger

and the resulting scenarios of use to the various communities that will build on this effort over the coming years.

Major Points

1. As the authors say, this is a proposal for a set of emission and land use scenarios that IAM teams are invited to develop. These will ultimately form the emission and land-use scenarios to be used by ESMs in CMIP7. I have two linked concerns with this.

First, while the different scenarios are well well-motivated and well described qualitatively, there appears to be only weak geophysical constraints on them. For some scenarios there are relatively clear constraints (such as the magnitude and timing of maximum warming in an overshoot scenario and the target long-term temperature post-overshoot). But, for others there does not appear to be clear geophysical constraints (e.g. global warming bounds, global mean radiative forcing bounds) for the IAM teams to aim for. Figure 1 gives one example of some geophysical outcomes for the proposed example scenarios Should the results there be taken as constraints by the IAM teams? In CMIP6 (SSPs) and CMIP5 (RCPs) the main geophysical constraint was global mean top of atmosphere (TOA) radiative forcing anomaly at 2100. So, I wonder whether a clear set of geophysical guardrails (upper and lower targets at certain time points) need to be provided for each scenario? There are numerous places in the paper with comments such as "*IAM teams should explore measures that minimize trade-offs and exploit synergies*" or "*IAM teams are encouraged to explore VLLO emission trends under different equity assumptions*". Isn't there a risk that the resulting IAM scenarios span too broad a range of geophysical outcomes?

We have edited Table 1 to describe the "Primary design characteristics" for the scenarios, which specifies the emissions or temperature outcomes that are desired for each. The low scenarios are defined by their intended temperature outcomes, while the medium and high scenarios are defined either by their assumed policies (Medium scenario) or their emissions characteristics.

Linked to my first concern, it is not clear what the decision-making process is for deciding which of a potentially large number of different IAM scenarios produced for each proposed scenario, will ultimately be used in CMIP7. ESMs will want one emission and land use data set per ScenarioMIP scenario. Given the relatively weak geophysical constraints on the different scenarios and the need for one "winner" a clearer

explanation of the decision-making process for selecting the "winner" for each scenario would be helpful.

We have described the marker scenario selection process in the introduction. There will be papers from the IAM community going into the details and technical considerations of this.

Linked to both concerns, I see at (Lines 137-139) *"the IAM scenarios based on this proposal are developed in the period Sept 2024 to summer 20205, so climate model simulations can start after summer 2025."* I note the word "*after*" summer 2025. Is the aim that the ESM-ready scenarios will be available by "summer 2025" or just the mix of IAM scenario submissions? If the ESM-ready scenarios, this seems ambitious but if it is the aim then I think some more guidance on (i) geophysical constraints per scenario and (ii) the decision-making processes to go from; numerous IAM scenario per scenario pathway to the single data set to be used by ESMs, needs to be included.

The text has been edited to update the timeline to the current status of the process, indicating that IAM marker scenarios will be finalized by the end of 2025.

Minor points:

1. Why was 2100 chosen as the end data for the IAM-based scenario? Previously IAMs have generated scenarios for 85 (CMIP6) and ~100 years. This seems to be reduced now to ~79 years. I am very supportive of the scenario extension plans, and understand the IAM scenarios become increasingly less well-founded with time into the future, but based on past CMIP cycles, I wonder why 2125 was not chosen as the end date?

   We fully agree that this would be desirable, and in the early stages of the ScenarioMIP planning process, there was ambition to explore a 2150 end date for IAM simulations. However, achieving this in practice requires assembling a large library of input data which defines time-dependent boundary conditions - for example defining scenario-dependent technological cost evolution, population growth, GDP growth at country-level resolution. Unfortunately, the tight timescale of the ScenarioMIP process precluded these updates/extensions from being finalised in time for ScenarioMIP CMIP7, but it remains an ambition for future scenario development and climate model intercomparisons.

2. Lines 191-193 talks about the CMIP6 emission scenarios being linked to different socio-economic futures (SSPs). At the time this was deemed to be beneficial for use of projections and scenarios linked through to societal impacts. So will the final CMIP7 scenarios (those chosen as the single representative emission scenario) also be linked to an assumed socoi-economic (SSP) future?

   In hindsight, the naming choice of CMIP6 has been a failure in communication, because there should be no one-to-one mapping of SSPs to RCPs. So in this phase we try not to emphasize the linkage as before. Individual scenarios will be by necessity produced using specific SSP assumptions and we will describe that, but the aim is to decouple SSPs from the emission outcomes as much as possible, since multiple SSPs can be used to develop the same RCP trajectory, and a single SSP can be used to develop multiple RCP trajectories. Hence the choice of names, this time around.

3. Lines 267-269 the authors say: "*Finally, it should be noted that models running emission-driven simulations can have different temperatures and concentration levels in the start year of the experiments. The ESM teams are strongly encouraged to keep deviations from observations in the historical period to a minimum*." I assume this is encouraging ESMs to try and have an accurate simulation of historical temperatures and CO2. This they generally try to do, without overtly tuning to every up and down in the historical records. As this paper is about the IAM scenarios I find this comment somewhat out of place here and I recommend dropping it.

   Thanks for this point. We agree that this text presented an unrealistic request for model development teams.

   We have modified the text as follows: "Running historical simulations in an emission-driven configuration presents a challenge, as a model's simulated CO_2 concentration can drift from observations. However, this approach is a deliberate and core feature of the CMIP7 experimental design. Rather than take measures to artificially correct this bias, we encourage modeling groups to quantify CO_2 concentration evolution as a key diagnostic of model performance and carbon cycle sensitivity. To facilitate this analysis, the ScenarioMIP protocol includes at least one parallel concentration-driven simulations specifically to allow for a clean diagnosis of the impact of these emergent carbon cycle feedbacks on climate outcomes."

4. Linked to my main point above, should IAM teams treat Figure 1 as a “target” for their scenarios? This is not really made clear. I also think global mean TOA radiation anomaly would be helpful to add to Figure 1.

   We have edited the text to indicate that Figure 1 is based on preliminary versions of the IAM marker scenarios (the figure is not a target).

5. With respect to the high emission scenario (H): Given recent geopolitical developments (e.g. US leaving the Paris Agreement), which of course happened after this paper was written, is there now a need for an increased sampling of the medium to high emission scenario space?

   The H scenario was already looking at possible reversal of trends for its justification. A HL scenario has now been added to the design, which increases coverage of the high end of the range.

6. Lines 512-513 discusses a so-called “competition scenario” that sounds like a particular SSP world, which again begs the question as to whether a range of SSPs will be represented in the resulting scenarios used in CMIP7 or whether this aspect of the scenario development is being abandoned?

   The text has been edited to give an up-to-date description of the scenarios, their basis in SSPs (all IAM markers are based on SSPs), and the SSPs underlying particular scenarios.

7. Text at lines 686 to 694, discussing options for the VLLO scenario, again reads to me as though there is a large amount of flexibility for what the resulting scenario is made up of. Without clear geophysical constraints I struggle to see what the decision process will be for selecting the single IAM scenario to be used by the CMIP7 ESMs.

   With the additional requirement of enough differences between the scenarios and plausibility constraints, the flexibility is somewhat bounded. Still, we asked several IAM teams to elaborate this scenario and finally chose the scenario that fitted the description best - also vis-a-vis the other ScenarioMIP scenarios. This process will be described in detail in the subsequent IAM papers.

8. Scenario extensions: These are important. It would be helpful to have a timeline for the availability of these as well.

The scenario extensions will be available at the same time as the scenarios themselves.

9. With reference to table 6. Is there a target resolution for the gridded BECCS and Afforestation data sets? And how is the DACCS data set delivered in terms of spatial distribution?

   We do not plan to deliver DACCS information to ESMs as part of ScenarioMIP. BioEnergy crops and afforestation will be passed through the LUH2 pipeline, and will be outputted at the native ¼ degree resolution of LUH2. Sequestration information for BECCS and DACCS will not be part of the ScenarioMIP dataset - and dedicated efforts like CDRMIP will work to extend the ScenarioMIP land use datasets to include gridded capture and sequestration products.

10. A lot of what comes under Discussion and Conclusions is really discussing things that would be nice to do in the future and could be removed. For example Lines 977 to 982 discusses high emission scenarios and high climate sensitivity models. This doesn't really seem to be part of this paper on emission scenarios. This is also the case for e.g. lines 990 to 994 and lines 1021 to 1025. I recommend shortening the text discussing future work here and instead more clearly summarize the main recommendations from the paper to give IAM teams a concise summary of what is aimed for and by when.

    The ScenarioMIP activity, and therefore the paper, targets a diverse audience, and we think it is important to address here many of the scientific future directions that could build upon this phase and facilitate the next, beyond the focus on this phase's IAM work. Some of the research efforts we encourage, if successful, could also alleviate criticisms we have been target of, and probably we will continue to be, about some aspects of the design that we are constrained to by necessity, i.e., lack of resources or time, in this phase of ScenarioMIP. That said, we have eliminated that bullet point about climate sensitivity and high scenarios, which was opaque and, we agree, did not add anything significant to the discussion.

David Huard (CC6):

Thanks for the paper, really appreciated. My comments below are from the perspective of a climate services provider. To give a bit of background, my office interacts daily with practitioners from multiple fields, looking to integrate climate projections into their work. Those practitioners are not climate experts, yet have to make decisions about climate adaptation options, and need clear guidance from the climate science

community on the use of scenarios. As climate service providers, our job is to translate the science being done in the research community into actionable information. ScenarioMIP is our main source of information, so the design decisions outlined in the paper will be critical to our work for the next 10 years. We have a strong vested interest in the experimental protocol of ScenarioMIP being understandable to a wide public.

Box 1

I disagree with Box 1' line of reasoning. The first paragraph mentions that the question under discussion is the "likelihood of a scenario approximating reality for a number of key output variables". But then the second paragraph states that for a scenario to be plausible, it has to be feasible based on the five dimensions of feasibility, which arguably cast a much wider net than "key output variables". I imagine a scenario could have an unfeasible demography with unfeasible technologies, that combined together would yield feasible emissions. I would argue that, from the ScenarioMIP perspective, the only dimensions that matter are the variables used as forcings for climate models, and assessments of plausibility should only look at those.

We have edited the text to make this point more clearly. It is true that ultimately what is most important for ScenarioMIP is the plausibility of the forcings provided to ESMs (emissions, land use, in some cases concentrations). However, to judge the plausibility of those outcomes, one needs to judge the plausibility of the conditions that produce them. For that reason, the dimensions of feasibility are relevant to judging the plausibility of the scenario outcomes for ESM forcings.

Stylized Pathways (L726)

The paper justifies the decision to use idealized extensions beyond 2100 by the "uncertainties that increasingly affect the socio-economic drivers of these trajectories". This feels arbitrary, and I believe the 2100 switch to stylized pathways is deeply problematic for multiple reasons.

> It breaks the first design principle outlined in the introduction: "internally consistent socio-economic and technological scenarios";
> It will damage the credibility of all climate impact assessments post-2100. Just imagine yourself explaining to a corporate board or elected officials that your climate risk analysis is based on a "stylized emission pathway";

> It creates a break in the provenance of climate projections. All regional, national and international web portals, including the IPCC Interactive Atlas, will have to explain why in 2100 there are abrupt and unrealistic changes in emissions. Communicating climate change projections is hard enough already, please don't make it harder by creating artificial breaks in the series;
> If modeling centers are also skeptical of these stylized pathways, they're less likely to run simulations, which will reduce ensemble size post-2100, hamper our ability to assess hazards, and lower the value of the whole exercise.

An alternative to stylized pathways would be to generate large Monte-Carlo IAM ensembles of emission pathways, and pick those that match best the intended stylized pathways. This would preserve internal consistency and would not muddy communications with end users.

Thank you for your thoughtful comments. We agree that the credibility and usability of the scenario extensions are paramount, especially for post-2100 climate impacts and risk assessment. Your concerns about the "deeply problematic" nature of abrupt, stylized transitions at 2100 have prompted a significant revision of our methodology.

While the absence of integrated assessment model simulations beyond 2100 necessitates a degree of stylization, we have taken several concrete steps to improve the internal consistency and plausibility of the extensions. Firstly, we will now implement smooth transition periods between the IAM-driven scenarios and the subsequent extensions. The pathways will be constructed to ensure a continuous derivative evolution of emissions, explicitly removing the abrupt changes that you rightly identified as a concern.

Furthermore, this principle of smooth transition will be applied across all forcing components. For non-CO_2 gases, we will use a set of documented, functional choices to ensure a smooth evolution from their pre-2100 behaviour to their long-term levels. For land use, we are replacing the previous step-function change at 2100 with a 50-year linear decline in AFOLU emissions to zero, which can be interpreted as a gradual phaseout of 21st-century forest management practices. We believe these changes directly address your critique and will provide a more credible and defensible set of long-term scenarios for the user community.

Minor suggestions

170: I would add "internally consistent" to the first design principle, as in line 180.

Agreed, we have edited accordingly

173: As discussed above, I object to this "design principle", which anyway feels more like an implementation detail than a principle.

Please see our reply to the comments on this issue just above.

Box 1: The first paragraph mixes two different ideas: the infinite number of potential futures, and the "dimensions" of each of those futures. It clarifies the latter by saying that we're mostly interested in key forcing variables, but does not address the former. One option would be to focus on the relative likelihoods of ScenarioMIP scenarios (a small, finite, set).

The number of potential futures is addressed in the current text by the reference to a scenario “lying in a given range”; i.e., not exactly matching a potential future outcome. While the likelihood of exactly matching the future is zero, the likelihood of approximating it is finite. Also note that we now discuss the relatively likelihood of scenarios in a new paragraph added just after Box 1.

451: Not sure I understand what the message from this table is, and whom it is intended for. I think a list of data that ScenarioMIP plans to make available, irrespective of whether it is intended for climate models or VIA, would suffice.

We beg to differ, as this paper is aimed at a much broader, diverse, multidisciplinary community, not just ESM modelers, so these distinctions are important to preserve.

498: Iea -> IEA

Corrected, thank you

947: I suggest to include among further research directions something about the relative likelihood of key forcing variables, e.g., something like "The IAM community is

encouraged to develop probabilistic assessments of ScenarioMIP' emission pathways to inform climate risk assessments and adaptation decision-making."

We have included a new paragraph in section 2.2 that discusses the probabilistic scenario literature and its relation to the ScenarioMIP approach of judging plausibility alone. We have also added a new item to the concluding section referred to here that encourages research on judging the relative likelihood of scenarios.

To expand on scenario likelihood, what decision makers are looking for can be translated as the (Bayesian) probability of the following statement being true: "Climate model drivers from scenario S are an accurate description of future conditions", for all S in the set of available scenarios. This is beyond the scope of this paper, but I hope the scientific community can finally tackle that question head-on. It's been 25 years since Moss and Schneider (2000) wrote: "We believe it is more rational for scientists debating the specifics of a topic in which they are acknowledged experts to provide their best estimates of probability distributions and possible outliers based on their assessment of the literature than to have users less expert in such topics make their own determinations."

Thanks again for your work on this, I can't overstate how important this is.

Mai M M Chim (CC7):
I would like to raise a discussion regarding volcanic forcing in the proposed CMIP7 ScenarioMIP framework. The current preprint does not address volcanic forcing, which I believe merits consideration given recent advances in our understanding of its importance for climate projections.

In CMIP6, ScenarioMIP simulations used a constant volcanic forcing based on the 1850-2014 historical average, which is biased by the under-recording of small-magnitude eruptions prior to 1978 (i.e., the start of the satellite era), and it does not account for eruptions larger in magnitude prior to 1850. Our recent study (paper accepted, preprint at https://www.researchsquare.com/article/rs-4938494/v1) demonstrates that volcanic forcing uncertainties contribute substantially to overall uncertainties in global mean surface air temperature projections – at a magnitude comparable to internal variability.

I recommend that CMIP7 ScenarioMIP incorporate an improved volcanic forcing representation that accounts for the climate uncertainty arising from future volcanic eruptions. This could be implemented through prescribing a constant volcanic forcing with magnitude equivalent to:

1. A historically-averaged mean that considers the missing sulfur dioxide flux from small-magnitude eruptions prior to 1978; or
2. The median stratospheric aerosol optical depth based on stochastic scenarios resampled from the latest ice-core and satellite volcanic emission records.

This approach improves the magnitude of volcanic forcing and the mean climate state in climate projections. I suggest modelling groups perform projections with constant volcanic forcing at the 5th and 95th percentiles of the stratospheric aerosol optical depth distribution of the stochastic scenarios to account for the climate uncertainty arising from volcanic eruptions.

I would also like to highlight that a constant volcanic forcing approach in ScenarioMIP does not consider the sporadic nature of volcanic eruptions. The use of stochastic volcanic forcings in climate models and projections is necessary to allow the assessment of the abrupt climatic changes caused by large-magnitude volcanic eruptions, and the associated climatic risks and socio-economic impacts.

We have added a new subsection to section 2.4 that describes natural forcings (solar and volcanic) in the ScenarioMIP scenarios. The CMIP7 forcing task team has developed a new historical reconstruction for stratospheric aerosol optical properties and volcanic sulfur emissions covering the period from January 1750 to December 2023 (Aubry et al., in prep). For the future scenarios beginning in 2025, a specific protocol will be followed to ensure a smooth transition and a consistent baseline for future projections.

The future volcanic forcing will involve a 9-year linear ramp-up from the historical values at the end of 2024 to a new background level. This background is defined as the mean stratospheric aerosol optical depth (SAOD) from the pre-industrial control (`piControl`) simulations, which is calculated over the 1850–2021 period. This target is higher than the mean SAOD of the more volcanically quiescent 2000-2020 period. The SAOD will ramp up until 2033 and then hold steady at this `piControl` climatological value for the remainder of the future simulations. Notably, the major eruption of Hunga Tonga-Hunga Ha'apai in 2022 is deliberately excluded from the future forcing scenarios to avoid its large and anomalous cooling signal from complicating the analysis of the anthropogenic warming signal in the scenario runs. Full details of the methodology and dataset are forthcoming (Aubry et al., in prep).

Gareth Jones (CC8):
I have a few comments I hope the authors will consider.

Lines 323-332

Figure 1 gives an illustration of the possible CO2 trajectories, and the global temperature response based on a FaIR model. But does it give a realistic estimate what would be expected from ESMs? Figure 7c in Sanderson et al 2024, shows an estimate of ESMs warming from CMIP6 models driven with CO2 emissions with a present day central 50% range of about 0.4C. This is much bigger than is suggested in Figure 1b.

I suggest that this is addressed, so that readers will have some awareness of how much larger temperature ranges will be in the CO2 emission driven experiments, compared to the CO2 concentration experiments.

We have better underlined the approximate nature of this figure, including the large uncertainty bands around the trajectories.

L268-269

How do ESM teams practically keep "deviations from observations in the historical period to a minimum", especially if extra radiative forcing uncertainty (from CO2 emission configurations) is present?

Thanks for this point. We agree that this presents an unrealistic request for model development teams.

We have replaced the text as follows:

“Running historical simulations in an emission-driven configuration presents a challenge, as a model's simulated CO_2 concentration can drift from observations. However, this approach is a deliberate and core feature of the CMIP7 experimental design. Rather than take measures to artificially correct this bias, we encourage modeling groups to quantify CO_2 concentration evolution as a key diagnostic of model performance and carbon cycle sensitivity. To facilitate this analysis, the ScenarioMIP protocol includes at least one parallel concentration-driven simulations specifically to allow for a clean diagnosis of the impact of these emergent carbon cycle feedbacks on climate outcomes.”

In O'Neill et al. (2016), there was a recommendation about the solar and volcanic forcing in the 21st century. What is the recommendation for CMIP7?

We have added a new subsection to section 2.4 that describes natural forcings (solar and volcanic) in the ScenarioMIP scenarios.

Solar Forcing

The solar forcing recommendations for CMIP7 are provided by the SOLARIS-HEPPA group. The historical forcing reconstruction covers the period up to December 31, 2023, and is extended through December 31, 2024. For future simulations from January 1, 2025, through 2299, the data uses an intermediate scenario from a stochastic ensemble of future solar activity, which is based on a surrogate analysis of cosmogenic isotope datasets. The complete recommendations and dataset construction are described in Funke et al. (2024) .

Funke, B., Dudok de Wit, T., Ermolli, I., Haberreiter, M., Kinnison, D., Marsh, D., Nesse, H., Seppälä, A., Sinnhuber, M., and Usoskin, I.: Towards the definition of a solar forcing dataset for CMIP7, Geosci. Model Dev., 17, 1217–1227, https://doi.org/10.5194/gmd-17-1217-2024, 2024.

Volcanic Forcing

The CMIP7 forcing task team has developed a new historical reconstruction for stratospheric aerosol optical properties and volcanic sulfur emissions covering the period from January 1750 to December 2023 (Aubry et al., in prep). For the future scenarios beginning in 2025, a specific protocol will be followed to ensure a smooth transition and a consistent baseline for future projections.

The future volcanic forcing will involve a 9-year linear ramp-up from the historical values at the end of 2024 to a new background level. This background is defined as the mean stratospheric aerosol optical depth (SAOD) from the pre-industrial control (`piControl`) simulations, which is calculated over the 1850–2021 period. This target is higher than the mean SAOD of the more volcanically quiescent 2000-2020 period. The SAOD will ramp up until 2033 and then hold steady at this `piControl` climatological value for the remainder of the future simulations. Notably, the major eruption of Hunga Tonga-Hunga Ha'apai in 2022 is deliberately excluded from the future forcing scenarios to avoid its large and anomalous cooling signal from complicating the analysis of the anthropogenic warming signal in the scenario runs. Full details of the methodology and dataset are forthcoming (Aubry et al., 2025).

Tejal Kanitkar (CC9):

I. It is incorrect to claim that the climate models "*are not very responsive to underlying assumptions on equity and justice*"- especially with respect to regional patterns of land use change (Box 2). In the case of global fossil fuel CO2 emissions, the corresponding warming levels and climate model outputs are dependent only on the emissions trajectory/forcing and not on the regional distribution of emission sources, and it follows that socio-economic assumptions underlying emissions trajectories can be decoupled from climate model outputs. However, this clearly does not hold for land use cover and land use change. Land Use scenarios for the corresponding emissions

scenarios proposed in this paper are intended to be generated using Integrated Assessment Models- and therefore involve explicit assumptions about regional incomes, population growth, food demand, food prices, production costs and trade. The range of land use scenarios used as inputs to Earth System Models- and therefore the range of socio-economic assumptions underlying them- will constrain the possible set of scenarios that can be generated subsequently for assessments in the upcoming IPCC AR7 cycle. Socio-economic assumptions already embedded in the land use change scenarios will foreclose various future possibilities of regional distribution of mitigation burdens, and therefore the question of equity is explicitly involved.

Therefore, we suggest that the ScenarioMIP for CMIP7:

1. State clearly the socio-economic assumptions underlying each land use scenario, and highlight implications for food demand, prices and costs and consequently food security, land availability for urbanisation, biodiversity conservation, agriculture, and energy crops.
2. Generate a baseline land use scenario consistent with present day land cover, that can be used as a control to isolate the effects of land use change by comparison with climate model outputs for other land use change scenarios.
3. Generate a set of land use scenarios that can be used in permutation with the proposed emissions trajectories, as opposed to fixing one specific land use scenario corresponding to each of the emissions trajectories.
4. The range of land use scenarios should necessarily include a set of scenarios with underlying assumptions of ensuring food security in the near term, and also meeting the Sustainable Development Goals.

We underline here that suggestions 1-4 will allow for the development of a wider range of scenarios. This is directly linked to allowing the possibility of constructing scenarios that take equity, justice, and sustainable development into consideration, as indicated by the approved outlines for AR7. Aspects of equity and sustainable development are directly embedded in land-use assumptions. The current proposal assumes that underlying assumptions on equity will not impact outcomes of Earth System Models. We submit that while even this claim must be examined (as discussed in the next set of points), it is undeniable that fixing land-use scenarios from the IAMs will inevitably narrow options for constructing equity-based emissions pathways.

Clearly, socio-economic assumptions in scenarios (including those related to equity) do impact fossil fuel emissions, land use outcomes as well as mitigation costs and the

impacts of climate change. The focus of ScenarioMIP is to provide a set of reference scenarios that explore climate change patterns using ESMs as a function of emission and land use scenarios. Here, so-far, multimodel studies have - at best - found only mild robus signals from underlying socio-economic assumptions (see Section 2.2). We already point out that it will be important to do further work on estabiiishing the exact relationships between land use and air pollution patterns and local climate change (see recommendations). It might be also be the interest of more specific MIPs to look into alternative land use and air pollution patterns (like LUMIP under CMIP6 did run alternative land use cases, very similar to what you describe above). Please also note that in the papers that will be published by the IAM community on the ScenarioMIP scenarios more details on the socio-economic assumptions will be provided. We disagree that the provision to ESMs of specific land use outcomes for each scenario will narrow options for considering alternative equity-based emissions pathways. Such alternative pathways can and should be developed in research of the WGIII research community and coupled with the climate outcomes produced by the ScenarioMIP scenarios.

II. Irrespective of whether ESMs are concentration-driven or emissions-driven, so long as land-use scenarios are generated from IAMs, the underlying range of socio-economic assumptions will, in principle, influence the range of ESM outputs for key climatic variables and carbon budgets. Whether the explicit representation of land processes in emission-driven runs will make outputs from climate models more sensitive to land use forcings and therefore to the underlying socioeconomic assumptions remains an open question.

The paper cites Tebaldi et al 2023 to state that “regional patterns of emissions of short lived forcers and land use… do not necessarily translate to a unique climate change pathway, since they have been shown to lack a robust regional climate signal in a multi-model context”. Tebaldi et al 2023 studies different land use change scenarios to show that the resulting ESM multi-model outputs do not give a clear regional signal in terms of changes in the occurrence of *extreme* climate events. This, in itself, is insufficient to claim that different regional patterns of land use give rise to equivalent climate forcings.

Modeling studies need to be done to clearly demonstrate the influence (or the lack thereof) of socio-economic assumptions on ESM outputs, rather than simply claiming apriori that ESM outputs are insulated from underlying socioeconomic assumptions.

In the introduction to the paper work is cited that has failed to find significant and coherent changes for average Temperature & Precipitation. Also, the paper does not consider grid-point-level 'far-in-the-tail' extremes. The extreme indices considered are designed to be impact-relevant but also do not ask much of ESMs in their extreme behavior, and the significance of the difference between experiments is computed using big regional averages, and choosing deliberately areas where aforestation/deforestation are most consistent and strong across ESMs. All these choices are trying to facilitate the emergence of the LUC signal, which remains elusive at best.

III. In considering the set of scenarios, we suggest the inclusion of a scenario that stabilizes at 1.7-18 deg. C. We understand that this depends on the sensitivity of ESM runs and if the ESMs can resolve the effects of this temperature difference significantly over internal climate variability. Since a separation of 0.25-0.3 deg. C was suggested for CMIP6, we suggest ScenarioMIP develop a scenario that stabilizes at 1.7-1.8 degrees by 2100.

This would be very policy relevant as overshoot (from 1.5 deg.C) scenarios imply a certain set of interventions required during the period of overshoot. On the other hand, the mitigation, adaptation, and other policy options available for a 1.7-1.8 deg. C stabilisation pathway would be significantly different as compared to a 1.5 deg. C overshoot pathway. The near-term implications of the responses required under both sets of scenarios (1.5 deg. C overshoot vs. 1.7-1.8 deg. C stabilisation) would also be considerably different, e.g. implications for food security if higher CDR is required in overshoot scenarios as compared to stabilisation scenarios. Instead of the current focus on overshoot pathways in the proposal, which in our view is disproportionate, the inclusion of a 1.7-1.8 deg. C stabilisation pathway will provide a more balanced and comprehensive view of future possibilities.

The current "Low (L)" scenario addresses the desired levels, does not present an overshoot, and is expected to end just below 2C. (Figure 1)

(Comments jointly submitted by Akhil Mythri, Tejal Kanitkar, and T. Jayaraman)

Fiona O'Connor (CC10):

> Thank you for putting together a proposal for ScenarioMIP's contribution to CMIP7 with the aim of underpinning contributions to the forthcoming Intergovernmental Panel on Climate Change assessment reports. In reading the discussion paper, we are grateful to you for the effort in putting this proposal together and the level of engagement with the community through workshops, reviews, and expert task groups. Likewise, we acknowledge the level of ambition required to develop illustrative pathways that meet different user and community needs.

It is good to see that the proposal captures a range of plausible future pathways, which is bound at one end by high levels of climate change corresponding to climate policy failure and at the other end by a number of pathways with stringent climate policies consistent with the Paris Agreement temperature targets. The range of proposed scenarios also includes two pathways that explore the implications of an overshoot which we fully support.

Thank you for your positive assessment.

Our main reason for commenting on your discussion paper is in relation to the assumptions surrounding air quality policies and how they will impact on emissions of short-lived climate forcers. While the pathways explore a range of plausible climate outcomes, they all assume rapid improvement in air quality and do not incorporate a diverse range in global and regional emission pathways for short-lived climate forcers. It is plausible that air quality policies may not progress as rapidly as assumed, and may even be reversed in some cases. Neglecting such possibilities risks missing possible climate outcomes. As you acknowledge, aerosol emission changes shape regional climate responses and will be one of the major drivers of climate change in the coming decades (e.g., Persad et al., 2022). Model results, e.g., those from the Regional Aerosol Model Intercomparison Project (RAMIP; Wilcox et al., 2023), indicate that aerosol emission changes are influencing global and regional temperatures and precipitation, including extremes. Likewise, future changes to air quality and exposure to poor air quality along with its health impacts are predominantly influenced by regional emissions rather than by global scale forcing and/or climate policies (e.g., Turnock et al., 2020). Specifically, the proposal under discussion does not consider how climate outcomes may depend on patterns of regional forcing and/or different forcing agents, and their importance for regional societal impacts and vulnerabilities. This limits the ability of AerChemMIP2 to define experiments to address

inequities in poor air quality exposure and associated regional climate responses due to regional diversity in short-lived climate forcer emissions.

We would therefore suggest that ScenarioMIP in CMIP7 considers diversifying the assumptions around air quality policies and consequences for short-lived climate forcer emissions in order to better represent the range of plausible future climate outcomes, and we would be open to discussing potential alternative pathways to the proposed scenarios. Some indication of how methane may differ across the pathways would also be useful, particularly given its implications for both climate and air quality and the potential for emissions-driven capability (e.g., Folberth et al., 2022) in CMIP7.

As a suggestion, we note that despite the different pathways, there is little separation in the global mean temperature pathways until 2060. This could be an opportunity for exploring alternative emissions pathways for short-lived climate forcers while not losing any of the diversity in the longer term global climate evolution.

Thank you for considering our suggestions and we look forward to hearing your response to our suggestions.

Fiona O'Connor, Bill Collins, Bjorn Samset, Steven Turnock, and Laura Wilcox

References:

Folberth, G. A., Staniaszek, Z., Archibald,A. T., Gedney, N., Griffiths, P. T., Jones,C. D., et al.: Description and evaluation of an emission-driven and fully coupled methane cycle in UKESM1, J. Adv. Modeling Earth Sys., 14, e2021MS002982. https://doi.org/10.1029/2021MS002982, 2022.

Persad, G. G., Samset, B. H., and Wilcox, L. J.: Aerosols must be included in climate risk assessments, Nature, 611, 662-664, 10.1038/d41586-022-03763-9, 2022.

Turnock, S. T., Allen, R. J., Andrews, M., Bauer, S. E., Deushi, M., Emmons, L., Good, P., Horowitz, L., John, J. G., Michou, M., Nabat, P., Naik, V., Neubauer, D., O'Connor, F. M., Olivié, D., Oshima, N., Schulz, M., Sellar, A., Shim, S., Takemura, T., Tilmes, S., Tsigaridis, K., Wu, T., and Zhang, J.: Historical and future changes in air pollutants from CMIP6 models, Atmos. Chem. Phys., 20, 14547–14579, https://doi.org/10.5194/acp-20-14547-2020, 2020.

Wilcox, L. J., Allen, R. J., Samset, B. H., Bollasina, M. A., Griffiths, P. T., Keeble, J., Lund, M. T., Makkonen, R., Merikanto, J., O'Donnell, D., Paynter, D. J., Persad, G. G., Rumbold, S. T.,

Takemura, T., Tsigaridis, K., Undorf, S., and Westervelt, D. M.: The Regional Aerosol Model Intercomparison Project (RAMIP), Geosci. Model Dev., 16, 4451–4479, https://doi.org/10.5194/gmd-16-4451-2023, 2023.

Since this comment, the trajectories with regard to air quality policies/air pollutants/SLCFs were discussed further with most of the researchers involved in this comment. Based on this, we have maintained the assumptions of improving air quality, and therefore overall consistent treatment, across the ScenarioMIP scenario set, of SLCFs. The reason was that it was found that the scenarios do cover a sufficiently large range in air pollution outcomes. We still believe that the best way to find out the exact impact of air pollution assumption is to test alternative patterns as variants of one another only with respect to SLCFs (as in the past happened with variants run by AerChemMIP of SSP1-2.6 and SSP3-7.0). If desired, the IAM groups plan to collaborate with AerChemMIP and produce these variants.

Anjali Sharma (CC11):

> The CMIP7 ScenarioMIP proposal presented here misses the opportunity to study and understand equitable pathways for climate action.
>
> 1) Firstly, the proposal refrains from exploring equity and justice concerns in its scenario framework because its main goal is to provide emissions data for ESMs to study climate outcomes. This exclusion is based on multi-model studies (Westervelt et al. (2020) and Tebaldi et al. (2023)) that find no robust differences in climate outcomes when considering different regional distributions of emissions from short-lived forcers and land use. As the ESM assessments are primarily sensitive to parameters like global emission levels and land use, and not much responsive to how these emissions are distributed across regions or sectors, authors argue that equity-focused distributional concerns are outside the scope of ScenarioMIP activities (see section 2.2; Box 2).
>
> Given this, it is surprising to find that later in the paper (see Box 3 and lines 1018-1020 on pg. 28), the authors note that, in fact, this assumption is supported by limited research and “further study of the degree to which variations in regional forcing due to short-lived climate forcers or land use could produce significantly different climate and impact outcomes would be useful” (lines 1018-1020). If the evidence on the impact of regional and sectoral distribution of emissions of short-lived forces and land use on

climate outcomes remains limited, then the reasoning of excluding equity and justice considerations in ScenarioMIP becomes questionable.

There are many aspects of scenario-related research that will benefit from further work; we list several in the concluding section. We don't believe it is inconsistent to establish the experimental design for ScenarioMIP based on the best current understanding of climate sensitivity to regional forcing, while also encouraging further research on this topic.

2) Secondly, the proposal misses the opportunity to bring in equity and justice discussions, which could be brought in even when discussing scenarios that result in the same forcing overall. The qualitative scenarios discussed in section 3 usually suggest the same policy and technology trends globally, without making any distinctions between developed and developing countries. However, there could be scenarios where emission trajectories between the two groups could diverge; for example, an increase in emissions in developing countries in the short- to medium-term, whereas a steep decline in developed country emissions.

For example, authors state that the medium emission scenario (M), which assumes the continuation of current policy, can be used to answer questions like, "What are the future physical, socio-economic, and ecological risks implied by current levels of climate change policy?" (lines 540-543). However, there could be a medium emissions scenario where the emissions increase in developing countries to meet their needs for better living standards, which is partly compensated by strengthened mitigation efforts beyond the current policy pledges in developed countries. Similar framing could be suggested in other scenarios as well.
Authors note that it would be interesting to explore medium (M) and medium-low (ML) scenarios "under a wider set of socio-economic assumptions (lines 564-565). This again presents an opportunity to explore differentiated assumptions on the basis of equity. This could be an important breakaway from previous scenarios that have ended up perpetuating existing inequities in energy access and carbon space between developed and developing countries into the future.
For designing low emission scenarios, the authors emphasize accounting for feasibility and plausibility concerns regarding technology adoption. It would be critical to note here that such assessments aren't only informed by technological considerations, such as availability of technology and its cost. Even for the same technology, the actual cost of adoption could be much higher in developing countries compared to developed

countries due to the persistent inequities in climate finance. This creates gaps between the feasible levels of technology adoption between developed and developing countries, even when both have committed to stringent climate action.
The authors mention that they remain 'agnostic' to the socio-economic pathways that are used in producing the emissions and land-use pathways (lines 195-195 on pg. 5). However, the qualitative scenario design is largely discussed within the SSP or SPP-like framework, especially for the mitigation levels discussed in low emissions scenarios. If the scenarios are largely produced within the SSP framework, it limits the opportunities to design and understand the climate outcomes of alternative socio-economic realities. However, given the tight timeline to produce scenarios based on this proposal (i.e., by summer 2025), how feasible would it be to incorporate alternative views in the scenario design?

We appreciate the several important research questions the reviewer identifies related to equity issues. We believe all of these are questions that should be pursued in future work, as we encourage in the concluding section of the paper containing a list of future research directions. We maintain that these questions are outside the scope of ScenarioMIP, which is focused on producing climate projections that can be used as input to such research and is forced, by computational constraints weighing on modeling centers, to choose a single trajectory per scenario to be used in ESM experiments..

3) Thirdly, it is curious to note that the authors are suggesting assessing SDGs along with the very low emission scenarios and not the other scenarios. The historical evidence, so far, suggests that many of the SDGs (such as poverty eradication, improved metrics of education and health) are correlated with increased energy use (and consequently, increased emissions). Achieving SDGs remains a key challenge for developing countries, and hence, it would be worthwhile to include them in consideration in other scenarios as well. At least, it would be useful to justify why these are included for low-emission scenarios only.

Our mention of the SDGs in relation to the VLLO (now VL) scenario is not meant to imply that SDGs are unimportant in other scenarios. Rather, it is meant to provide a further constraint on this scenario given the potential for rapid emissions reductions to impact development concerns. We have edited the text describing the VL scenario to make this clear, and note that we also point to future research needed on sustainable development pathways in the concluding section.

Oliver Geden (by email)

In general

Given the prominence of the term “overshoot”, you should clarify early on, even if only in a footnote, what you mean by it. I guess you are relying on the AR6 understanding of “exceeding a GWL and returning to it”, but I wouldn’t count on other scientists having exactly that trajectory in mind (AR6 WGII used this definition but seems to talk only about 1.5C exceedance in its SPM, under the header “overshoot”). See also Reisinger et al. 2025 https://doi.org/10.1146/annurev-environ-111523-102029 (sorry for the self-promotion)

At the first instance when the term is used in the text without an explicit reference to a temperature level we have added a parenthetical: “(intended here as temperature trajectories that exceed a global warming level of interest for a limited time and later experience a decline due to declining GHG concentrations from negative emissions) “. Please also note that in the current version the naming is going to avoid that term altogether.

217

„mitigation and CDR technologies“ makes it sound as if CDR is not mitigation (while CDR is part of mitigation, i.e. reducing emissions and enhancing removals, as per UNFCCC). I guess you wanted to say “emissions reductions and …” (not sure if you need the “technologies” after CDR…)

We have edited accordingly.

295/296

What does it mean that you take “no position on Paris consistency of the low emission scenarios”? I have an idea but I think you should explain that. And why are you then talking about “informing policies consistent with the Paris Agreement” soon after? In general, I think there’s no need to touch this consistency issue in a ScenarioMIP paper since it is an issue for legal scholars (and among them you will find widely different interpretations)

We agree that this reference to consistency is confusing and unnecessary, so we have deleted this sentence.

304

"Paris Temperature Goals": in legal terms, there aren't temperature goals (*plural*) in the Paris Agreement, there's only one "Long-Term Temperature Goal", ranging from well below 2C to 1.5C (Rajamani/Werksman 2018: https://royalsocietypublishing.org/doi/10.1098/rsta.2016.0458)

We have changed from plural to singular "goal"

311

Why "Carbon Dioxide Removal" here? All these scenarios include gross CDR, so this is probably about "net-negative emissions"? Looks like you're using the latter in table 1 already

Good point, we edited to refer to net negative emissions.

351-352

Why does the full name of VLLO include a "with" while the full name of VLHO includes a "after". Is there a difference in meaning intended? If so, it would be wise to explain it. If not, it would be wise to use the exact same term (I don't have a strong position which one is better). If you unnecessarily use two different terms, people will mix them up all the time

The longer scenario names have now been changed and simplified to Very Low (VL, used to be VLLO) and Low-to-Negative (LN, used to be VLHO).

389-390

I don't disagree with the box but the "as it is covered in a separate MIP" should maybe go out because there are also other MIPs where this is not an argument not to include an approach or analytical dimension in these experiments, e.g. CDRmip

Correct. We have reformulated the text somewhat - to indicate that SRM is covered in another MIP (so providing a reference; not an argument not to cover it).

538-552 & 567/568

Hard to understand how you would want “current policies” to be operationalized, and you should clarify that, at least in the sense that such a decision would need to be made, and by whom and when. May just via the IAM scenario input? Rogelj et al. 2023 operate with 5 different cases, and I guess you would need to decide on one to make it work – or is it just their “Case A”? Later you refer to the CurPol scenario assumptions from AR6 WGIII, these include pledges as well, so that would lean towards “Case E”? Regardless of the exact definition, a relevant issue is of course the cut-off, especially because of recent changes in the US, and in case you include an NDC component, then how (not) to consider emerging 2035 NDCs?

There are indeed several cases on how to extend current policies - covering even more dimensions than considered by Rogelj et al (who mostly consider the credibility of reduction claims). In ScenarioMIP, we focus mostly on the current policies - and do not include the NDCs or the net-zero targets. The extension is done by assuming that policies are kept in place after their due date - but not strengthened (so “frozen”). This will be discussed in more detail in the relevant IAM papers.

550/551

Better say “the NDCs for 2030” (or something with “aggregated”, because it is not about the individual NDCs)

Edited this to “the aggregated NDCs expressed for 2030”

572-582

The scenario description could be clearer in saying explicitly that 1.5C is reached only after 2100 again, like in lines 304/305.

Added “aiming at the 1.5C level on a multi-century time scale.”

576/577

What does “impacts of overshoot” mean here? And why only after 2100? Probably “impact during overshoot” would be better, or “impacts following from overshoot” (covering the irreversible ones as well). See again Reisinger et al. 2025

Edited as in “Moreover, after 2100 AD the scenario could be used to explore the impacts during and after the temperature overshoot.”

662

“the very low ScenarioMIP pathway” is confusing here, because you are talking about this pathway right here. Maybe you mean the VLLO?

Edited, now using the scenarios’ names.

668

“Relevance to the Paris Agreement” is a stub.

We have elaborated: “The scenario temperature outcome aims to be consistent with the Paris goal and useful, as part of the triplet of low scenarios, to compare outcomes within that narrow range.“

698/699

Maybe better “the C3 category of IPCC AR6 WGIII”. The text in parentheses could probably be cut, because C3 was not explicitly declared to be an interpretation of the upper limit of the “1.5C to well below 2C” range. But even if you would want to make this connection, then better to bring in the Paris language somewhere and say that you interpret the “likely below 2C” in modelling terms as “well below 2C” in legal-political terms, because the latter community usually doesn’t get that and sometimes likes to complain about C3 in the sense that it stays only “below”. But again, maybe better to stay out of this completely

We have edited the reference to IPCC and dropped the text in parentheses as suggested by the reviewer. We have added a separate sentence about broad relevance to the Paris Agreement without getting into legal terms.

701/702

What does the sentence “Before 2070, some CDR use might compensate for hard-to-abate emissions sectors” actually mean? First of all, there are no “hard-to-abate sectors”. You don’t abate sectors, you abate emissions from some processes in certain sectors. But even beyond that, the meaning is not clear. Is this simply about that “some CDR is used”? But that is always the case. There’s no scenario without gross afforestation/reforestation, and in getting to net-zero by 2070, you also need to ramp-up the CDR part to the levels needed in the year of net-zero. But beyond that, does the “compensate for” carry any specific meaning, in the sense of model choices to bring sector after sector to net-zero via CDR ramp-up in respective sectors?

We have edited the text to refer to hard-to-abate emissions, rather than sectors, and also to substitute “offset” for “compensate for.” We leave the other details of this scenario to the IAM implementation.

723/724

In the L-ext/Storyline cell, shouldn’t it be “net-negative CO2 emissions" instead of (gross) “removals”? And in the VLHO-ext/Soryline cell, “net-negative Co2 emissions” instead of just “negative emissions” (the latter only if you want to use the same system as in AR6 WGIII, admitting that we failed to keep governments on this track during AR7 outline agreement, but that can be corrected later)

Thanks for this, the table has been revised

735

Again, “negative CO2 emissions” alone is suboptimal, I guess “net-negative CO2 emissions” is what’s meant here.

Again, thanks - well noted - we have revised

751

Although you indicate earlier that you don’t take a position on the “Paris consistency” of certain low scenarios, here you implicitly do it again. So VLHO is not Paris-consistent? Is this a known fact? Why does this need any mention in a ScenarioMIP article?

In response to the previous comment referred to here, we have deleted the sentence on not taking a position on Paris consistency. Regarding this comment, we have edited the text to make a broader reference to relevance to the Paris Agreement goals. Mentioning the Paris Agreement is logical since the design of the IAM scenarios was done explicitly in terms of 1.5 C and 2 C outcomes, as is now made clearer in a modified version of Table 1.

812

Again, “net-negative CO2 emissions” would be preferable here

The text has been changed to net-negative emissions

840

I guess it should be “removals” instead of “reductions”?

Edited

869

Here again, “emissions/reductions”. The latter is either “removals”, or I do not understand what you are getting at here, and I would probably not be the only one

Edited

936/937

“Storage of Carbon from CCS” is not an “AFOLU process”, maybe the categorization in left column needs a rethink

We relabelled this column as “BECCS sub-processes” to better describe the content of that column.

945/946

The text in the DACCS row is hard to understand. It may need a better explanation in section 5.1. What I’m confused about is the term “net emissions associated with DACCS”. This sounds like you are aiming for some kind of lifecycle emissions approach, but I may be wrong. Wouldn’t calculating net DACCS emissions/removals (if I understand correctly what is meant here) potentially lead to double counting (you have the energy emissions already elsewhere, as well as industrial emissions for the material build-up, maybe even transport) and wouldn’t it be inconsistent with other approaches where emissions & removals in supply chains are reported separately (like zero emissions for wind energy in the energy sector, but emissions from associated steel and cement production under industry?). Maybe I completely misunderstand what you are aiming at and why, but then I’m probably not the only one and it would be better to expand (here or in 5.1)

The intention here was not to imply a life-cycle emissions approach but rather to indicate that net emissions associated with DACCS would be provided as an independent stream of data to ESMs as well as included within total net emissions provided to ESMs, to facilitate attribution studies etc. However the new plan for CMIP7 ScenarioMIP is just to provide the regional net DACCS emissions as part of the total net emissions, so this part of the table will now be rephrased to: Regional net emissions associated with DACCS included as part of the total emissions provided to ESMs.