Reply to reviewer comments for RC1
https://doi.org/10.5194/egusphere-2025-5669-RC1

Dear Jean-Francois Lamarque, thank you for your review of the AerChemMIP2 experiment protocol. Your thoughts are helpful in further improving the written presentation of the protocol. Our replies are in blue beneath the comments in black.

This paper serves as the description, justification and documentation of the simulations intended to be part of the AerChemMIP Phase 2 in CMIP7. As such, it is a very descriptive paper and will indeed provide modeling centers the information they need to perform the simulations. Most of my comments aim at clarifying the text in terms of its simulation details, but also in its justification of choices. These should be addressed by the authors.

Line 87: this is a very large data request. Is there a clear understanding and documentation that this is useful? Maybe from AerChemMIP?
Without doubt AerChemMIP2 requests many individual variables owing to the aim of the MIP to address emissions and their effects on the level of individual perturbations for gases and particles. The compilation of the data request from AerChemMIP2 was a carefully designed and thoughtful community effort. Justifications were defined for each variable group, documented and published via AirTable. The release of the AFT data request for CMIP7 involved several intermediate steps with community consultation, coordinated via the CMIP data request task team (https://wcrp-cmip.org/cmip-phases/cmip7/cmip7-data-request/, last access 6 February 2026). Data for driving Chemical Transport Models (CTMs) is ranked with a lower priority (not yet included in the data request) and will for instance be used by GEOS-Chem for transient simulations. While the CTMs require additional data from AFT experiments, that output for driving CTMs helps to reduce the overall computational burden for modelling centers. The CTM simulations can for instance focus on hydrogen without the need to perform additional fully coupled Earth system simulations.

We add the reference to the atmosphere themed article for the CMIP data request including the AerChemMIP2 data in our revised text: „The potential implication of a hydrogen economy can be studied with Chemical Transport Models (CTMs) driven by output from selected CMIP7 model experiments that have been included in the full CMIP data request for atmospheric variables (Dingley et al., 2025).“

Dingley, ., Anstey, J. A., Abalos, M., Abraham, C., Bergman, T., Bock, L., Fiddes, S., Hassler, B., Kramer, R. J., Luo, F., O'Connor, F. M., Šácha, P., Simpson, I. R., Wilcox, L. J., and Zelinka, M. D.: CMIP7 Data Request: Atmosphere Priorities and Opportunities, EGUsphere [preprint], https://doi.org/10.5194/egusphere-2025-3189, 2025.

Line 89: why just “idea”?
Changed from „with the idea of creating multi-purpose experiments“ to „for creating multi-purpose experiments“

Line 97: what is CACTI?
The Composition Air quality Climate inTeractions Initiative (CACTI) is an umbrella for scientific exchange for MIPs with similar and complementary interests. It has evolved from earlier exchange between PDRMIP, RFMIP and AerChemMIP and the organisation of joint workshops. CACTI includes new MIPs, maintains connections with communities within and outside of CMIP, and organises annual hybrid CACTI workshops generously funded through the community. We now explicitly mention this process and link to our perspective article for additional information on CACTI: „via online meetings and workshop of the Composition Air quality Climate inTeractions Initiative (CACTI, Fiedler et al., 2024)“.

Fiedler, S., Naik, V., O'Connor, F. M., Smith, C. J., Griffiths, P., Kramer, R. J., Takemura, T., Allen, R. J., Im, U., Kasoar, M., Modak, A., Turnock, S., Voulgarakis, A., Watson-Parris, D., Westervelt, D. M., Wilcox, L. J., Zhao, A., Collins, W. J., Schulz, M., Myhre, G., and Forster, P. M.: Interactions between atmospheric composition and climate change – progress in understanding and future opportunities from AerChemMIP, PDRMIP, and RFMIP, Geosci. Model Dev., 17, 2387–2417, https://doi.org/10.5194/gmd-17-2387-2024, 2024.

Line 160: that is not a clear demarcation between chemistry packages. Even stratospheric chemistry focused packages have a representation of tropospheric chemistry (like methane oxidation and ozone production). Complexity in HC chemistry is probably a better split
Indeed, models have different capabilities for representing chemistry that may include reactions throughout the stratosphere and troposphere, or parts of it. To keep it simple, we remove the demarcation of tropospheric ($CHEM^T$) and stratospheric ($CHEM^S$) chemistry in the manuscript.

Line 161: why "as much capability as possible"? Doesn't that make comparison between model result more difficult? This should be justified
We add the following justification in the revised manuscript: „This diversity in model complexity is deliberately chosen, since too few model contributions in AerChemMIP was challenging for interpreting results and models with intermediate complexity are useful for transient experiments (Griffiths et al., 2025)."

Lines 170-172: this is very unclear. What is the point of using a version that has not done the DECK? It seems like you're enabling using an unconstrained version that could be giving very different results (on say ECS) and not be documented

Thanks for pointing out this lack of clarity. Is was not meant as an option when we write: „Irrespective of the level of model complexity chosen for participation in AerChemMIP2, it is desirable to maintain consistency with the setup used in the DECK and ScenarioMIP-CMIP7, whose experiments in some instances act as control experiments for AerChemMIP2." We change it to „we request that the model complexity is consistent with that used in the DECK and ScenarioMIP-CMIP7" in the revised manuscript. We also add in Section 3.2: „All experiments *histSST-piX* are requested to be consistent with historical experiments of CMIP7."

Lines 212-214: Shouldn't that be consistent with the way the model is going to be run for the historical simulations? Is the idea here that all changes in biomass-burning emissions are anthropogenic? I don't think that's the case, especially in the boreal regions.
We state: „The climatology is based on the last thirty years from the model's own PI fully coupled control experiment for *piClim-X* and the model's own fully-coupled historical experiment for *pdClim-X* experiments in AerChemMIP2", and we encourage these model experiments with interactive fires switched on (see also Section 3.2.5). We do not attempt to separate anthropogenic from naturally induced fires in all models contributing AerChemMIP2 experiments for ERF estimates, but we indicate how we think it could be done for some specific model capabilities to provide some guidance. We clarify this point in the revised manuscript as follows: „Not all biomass burning emissions have an anthropogenic origin. (…) For instance, in models which interactively simulate agricultural burning alongside unmanaged wildfire emissions, a separation between natural and anthropogenic emissions might not be possible."

Line 224: equilibrium in what? This has to be better defined
We revise it to: „We recommend to run the spinup until the simulated radiation balance at the top-of-the atmosphere and the chemistry state reach an equilibrium, i.e., trends across annual values (if any) become small. Some simulations with CMIP6 models required for instance 10 years for ozone depletion to reach an equilibrium after the concentrations of ODSs were changed from PI to PD levels (Figure S2 in Collins et al., 2026)."

Collins, W. J., Daniel, J. S., Chipperfield, M. P., Cussac, M., Deushi, M., Faluvegi, G., Griffiths, P., Hodnebrog, Ø., Horowitz, L. W., Keeble, J., Kinnison, D., Naik, V., O'Connor, F. M., Shindell, D., Tilmes, S., Tsigaridis, K., Wang, Z., and Weber, J.: Indirect climate forcing from ozone depleting substances, EGUsphere [preprint], https://doi.org/10.5194/egusphere-2025-6033, 2026.

Line 264: if the "weak dependency" has already been documented, why should this setup be part of the proposed simulations?
The cited study uses a CMIP7-like model without interactive aerosol and chemistry. The requested AerChemMIP2 experiment allows to understand the influence of interactive aerosol and chemistry, for which evidence from GFDL-ESM4.1 exists for anthropogenic aerosol ERF (Zhang et al., submitted). We add in the revised manuscript: „Models without interactive chemistry and aerosols, as often used for CMIP experiments, however, might only show a weak dependency on

the base state, e.g., for anthropogenic aerosol ERF (Fiedler et al., 2025; White et al., 2025).“, with two recently published articles now added.

Fiedler, S., Pham, T. V., Schlund, M., Wahl, S., Sudarchikova, N., Bischof, S., & Hoesly, R. M. (2025). First analysis of climate forcing and response to updated historical anthropogenic aerosol with the new CMIP7 model ICON XPP. Journal of Advances in Modeling Earth Systems, **17**, e2025MS005067. https://doi.org/10.1029/2025MS005067

White, K., D. Liu, and G. Persad, 2024: Absence of Aerosol Indirect Effect Dependence on Background Climate State in NCAR CESM2. *J. Climate*, **38**, 147–163, https://doi.org/10.1175/JCLI-D-23-0755.1.

Zhang, S., V. Naik, L. W. Horowitz, and D. Paynter: The Dependence of Emissions-based Anthropogenic Aerosol ERFs on Background Composition and Meteorological States in the GFDL-ESM4.1 Model, submitted to J. Geophys. Research.

Section 3.1.3: it seems that this is not very well developed and specific requests should be made much clearer to be useful. In particular, documentation on the aerosol scheme (which is always a tricky one in the CMIP documentation) needs to be more complete than simple “interactive” or “prescribed”
Indeed, more documentation of the model setup would be useful. To better reflect that point in the text, we revise it as follows: „Output from kilometer-scale simulations from models with any implemented aerosol treatment would be welcome, including both models using interactive aerosol schemes and models with prescribed data, e.g., for aerosol concentrations or optical properties. We ask the modelling centres to document their aerosol treatment in the metadata of the model output, e.g., with information on their aerosol treatment along with the reference for the prescribed aerosol data or the implemented aerosol parameterization schemes.“

Line 311: don’t most models have some basic H2 chemistry? But most likely not enough to do it properly? By the way are you asking for emission-based H2 simulations or surface concentration-based?
Most models do not have enough chemical interactions needed for an assessment of H2, with two known exceptions stated in the manuscript: „namely UKESM1.0 (Brown et al., 2025) and GFDL-AM4.1 (Paulot et al., 2021)“. We revise the text for clarity: „Although, policies for H_2 are currently not planned, implications of H_2 leakage for climate change are beginning to be quantified (Paulot et al., 2021, Brown et al., 2025, Chua et al., 2025). (…) Interested modelling centres are asked to contact the authors for coordinating the experimental setup in more detail, e.g., for using emissions or concentrations of H_2 for the experiment.“

Line 333: Based on AerChemMIP, how many Tier 3 simulations were actually performed by enough models to make them useful to be part of CMIP?
Griffith et al. (2025) give an overview about the data used from AerChemMIP phase one experiments. Tier 3 experiments were used for ERF and feedback estimates, both of which were also used in AR6 quite apart from several scientific publications. Specifically, 4 to 11 models contributed to Tier 3 experiments in AerChemMIP (Tab. 1, Griffith et al., 2025). One could say more model contribution would be better but some are certainly better than none. Having tiers in AerChemMIP2 is meant to help temporally planning the completion of experiments, and is not indicating importance or expected uptake by the community. We now explicitly comment on this point in section 5 as follows: „We encourage modelling centres to contribute to AerChemMIP2 experiments listed under all Tiers, which are meant as a help for temporally planning resources for completing experiments rather than misinterpreting tier 1 as being more important than tier 3 experiments. Indeed, tier 3 experiments from AerChemMIP phase one have for instance been cited similarly often as experiments in other tiers and informed IPCC AR6 thanks to contributions from many (Griffith et al., 2025).“

Line 350: What is the advantage of a SST+4K experiments over simply a future SST runs? I can see why it would be useful for CFMIP, not so much for AerChemMIP. Justify why this should be in AerChemMIP2.
The aim is to separate the contributions from warming on natural emission changes based on the preindustrial climatology (Section 3.1.6), an aspect that was not addressed in AerChemMIP.

Having the same experimental setup in several MIPs reduces the overall computational burden for modelling centres. We add: „Output from *piClim-p4K* can be used to isolate the response of emission changes to a temperature change and to eliminate any conflation between the biophysical and radiative effects of CO_2. The requested experiment piClim-p4K is the same as in RFMIP2 (Kramer et al., 2025) to help reduce the overall computational burden of modelling centres."

Kramer, R., Smith, C., and Andrews, T.: The Radiative Forcing Model Intercomparison Project (RFMIP2.0) for CMIP7, EGUsphere [preprint], https://doi.org/10.5194/egusphere-2025-4378, 2025.

Line 355: But, with that choice, the results will be a mix of models with and without biomass burning feedback? That will the intercomparison and attribution more complicated. Why is that the better approach?
Indeed, we now make this mix of models explicit at the start of the section: „In AerChemMIP2, we expect contributions from models with and without biomass burning feedbacks. The mix of representations of biomass burning across contributing models has the potential to better link to assessments of fires based on experiments in AFT." It might later for instance also be feasible for some models with biomass burning feedback to perform a sensitivity simulation to isolate the impact from the inclusion of the biomass burning feedback compared to prescribed fires expected in AFT experiments." To better account for the different model capabilities in AerChemMIP2 compared to AerChemMIP, we also request an experiment with doubled lightning flash rate that allows to investigate influences on atmospheric composition and radiative forcing that occur in addition to increased NO_x.

Line 480: what is the rationale for a linear increase in dust emissions? That seems highly subjective and unclear why that would be a useful experiment
It is an idealised future experiment for understanding what an increase in dust would imply for atmospheric composition and climate change. Hence we write here: „In addition to the historical experiments for dust, we request a future extension with a linear increase of dust aerosols in an overshoot scenario (*esm-scen7-vl-Dust*; see Section 3.3.1). Note that this future increase in dust is hypothetical in that there is currently no consensus on how dust will change in the future (Pu et al., 2018, Kok et al., 2023), although that gap in our understanding is known for at least two decades (Mahowald et al., 2003, Tegen et al., 2004).", and add in the scenario description: „The idealised design of experiment *esm-scen7-vl-Dust* allows us to investigate to what extent the climate response to dust forcing might be linear in a future scenario targeting net zero."

Lines 488-490: Dust is actually a critical component to the PI-control tuning. How would this adjustment be consistent with that tuning?
Indeed, thanks for pointing it out. We will provide community support for performing that experiment. A way to leave the preindustrial tuned dust untouched is the use of temporal scaling factors based on the historical dust data that will be applied to the model's own pre-industrial dust fields. The pre-industrial control simulation therefore will not necessarily have to be repeated unless model teams choose to do it.

Lines 532-533: what is the relevance of that sentence to AerChemMIP2?
We here link to FireMIP for additional experiments targeting fires that are not addressed in AerChemMIP2, and point to the new FireMIP protocol available as preprint (Li et al., 2025).

Li, F., Lawrence, D., Rogers, B., Burton, C., Huang, H., Jiang, Y., Kaiser, J., Kasoar, M., Lee, H., Leung, R., Nieradzik, L., Wang, A., Ward, D., Ce, L., Li, Y., Lin, Z., Voulgarakis, A., and Xue, Y.: The Fire Modeling Intercomparison Project (FireMIP) for CMIP7, EGUsphere [preprint], https://doi.org/10.5194/egusphere-2025-6115, 2025.

Line 579: what is the meaning of "treatment of SLCFs"?
We revise it for clarity: „and vary the future developments of SLCFs".

Line 585: aren't those all the SLCF emissions?
We clarify in the text: „(…) in which aerosols, O_3 and their precursor emissions are set to PD level while all other emissions follow the h scenario. (…) except for aerosols, O_3, and their precursors".

Please note that AerChemMIP2 follows the definition of short-lived climate forcers as in AR6 (see introduction).

Lines 603-604: is that the same linear extension described before? This is unclear
Yes, we now connect both sections: „(compare Section 3.2.4)“ and state that it is a hypothetical increase: „The experiment assumes a hypothetical increase in dust-aerosols over time with a rate similar to what has been seen for past decades. The experiment applies a linear increase to the dust emissions on the PD spatial pattern of sources. The future extension is part of the dust dataset listed for AerChemMIP2 and CMIP6plus via input datasets for Model Intercomparison Projects (ESGF input4MIPs, 2025).“

Lines 734-735: why mentioned that those “could be explored”? If they are not, then they shouldn’t be in this paper.
Changed to: „exist“.

Lines 741-742: this is way too general to be useful. What diagnostics? Isn’t this what AeroCom does already?
Agreed, details are not yet known and not essential for documenting the AerChemMIP2 experiments. We remove the two sentences in the revised text.

Line 762: even just one or two historical simulations is an enormous amount of data? Is that 3-hourly data? Who would run a CTM for the whole time-period?
The driving fields for CTMs are requested from single realisations with lower priority and will likely be limited to time slices (see also reply above). We add a citation of Dingley et al. (2025) for further information on the data request.

During the revision of the manuscript, we would also like to update the names of the scenario VLLO to VL, Figure 7 to be qualitatively rather than quantitatively to better reflect the preliminary nature of the scenario data, and clarify the nature of the illustrative scenario data in the acknowledgement.

Reply to reviewer comments for RC2
https://doi.org/10.5194/egusphere-2025-5669-RC2

Thank you for your review of the AerChemMIP2 experiment protocol. Your comments are helpful in further improving the written presentation. Our replies are in blue beneath the comments in black.

This paper presents the AerChemMIP Phase 2 simulation design for CMIP7, clearly outlining its scientific goals and experimental framework. It is exciting to see the new experimental designs in AerChemMIP2 compared to Phase 1 and the scientific opportunities they enable. The paper provides detailed and practical guidance for modeling centers and offers a clear overview for future users of the resulting multi-model dataset. I only have a few comments:

Line 249, how is the nitrate aerosol treatment expected to change in CMIP7 models? Could the authors provide more details?
In AerChemMIP, 2 models performed the piClim-NH3 experiment and 11 models piClim-SO2 and piClim-BC (Griffiths et al., 2025). We are aware of at least one additional model that can interactively simulated nitrate aerosols. We change the text to: „additional models with the capability of interactively simulating nitrate aerosols“

Line 277, the NMVOCs experiment is only designed for pdClim-X (table 2) but not for piClim-X (table1). The rationale for this choice is not clear and should be explained.
Thanks for spotting this. We remove NMVOCs in this line referring to both piClim-X and pdClim-X experiments. The choice for using pdClim-X experiments for NMVOCs is now explained in Section 3.1.5: „We choose here pdClim-X experiments for enabling direct comparison to observational data sets for present-day conditions.“

Line 345, the -2X experiments are designed for piClim but not for pdClim. Could the results be affected by state dependence if similar perturbations were applied under PD climatological conditions? Please consider adding a brief note discussing potential state-dependence issues here.
As you rightly point out, we have included the -2X experiments for the piClim experiment and not for pdClim. The piClim-2X experiments here give some traceability back to Phase 1. Given the non-linearity of aerosol-cloud interactions, for example, we may well expect the radiative effect of doubling aerosol emissions in an anthropogenically-perturbed present-day state to be different to that in a pre-industrial state. As much as we might have wanted to include pdClim-based doubling experiments to address state dependence, we had to strike a balance between that and the total number of experiments. However, with modelling centres setting up and running pdClim-control, there is potential for modelling centres to independently investigate such state dependence by running pdClim-2X experiments.

We've added the following to the manuscript: „The state dependence of the interactive emissions (if any) is not yet well understood. To what extent the interactive emissions change due to a simulated warmer climate state compared to the preindustrial can be addressed with the newly added *piClim-p4K* experiment, (…). Output from *piClim-p4K* can be used to isolate the response of emission changes to a temperature change and to eliminate any conflation between the biophysical and radiative effects of CO_2. “

Line 392/398, “Aerosol precursors” and “O3 precursors” are discussed as early in Table 1 but are not defined until this paragraph. I suggest defining these terms earlier in the paper. Also, it is somewhat confusing that NOx and VOCs are categorized only as O3 precursors but not as aerosol precursors, given that they also contribute to aerosol formation. Please clarify and justify this categorization. Relatedly, how are interactions between aerosol and O3 treated or separated in the AerChemMIP2 experimental design?
We now introduce *aer* in Section 3.1.1: „The experiment *piClim-aer* in AerChemMIP2 uses aerosols and aerosol precursor emissions of BC, organic carbon (OC), ammonia (NH_3), and sulphur dioxide (SO_2), and omits changes in anthropogenic NOx and VOC emissions. Although these latter species may act as aerosol precursors and influence oxidising capacity and secondary aerosol formation, they are omitted here to be consistent with the same experiment in

RFMIP2 (Kramer et al., 2025) and to maximise potential contributions from models with AER capabilities.“

Table 5, I suggest explicitly listing the reference experiments in the table (e.g., hist-piControl, esm-scen7-h, and esm-scen7-vllo). In addition, the reference for the –h and –vllo differs, but the same label (“Ref”) is used in the table, which may be misleading. Using distinct labels, as done in Table 8, would improve clarity. A similar issue also applies to Table 7.
Good point, we list the scenarios explicitly rather than ref in Table 5 and 7.

During the revision of the manuscript, we would also like to update the names of the scenario VLLO to VL, Figure 7 to be qualitatively rather than quantitatively to better reflect the preliminary nature of the scenario data, and clarify the nature of the illustrative scenario data in the acknowledgement.