Response to reviewer comments

We thank the Editor and Reviewer 3 for their constructive feedback on our manuscript. Our point-by-point responses to their comments are in blue below.

In addition to addressing their comments, we have made a small change to the protocol. We have realised that the *piSST* experiment was not needed, because the *piControl* climate (which *piSST* replicates) can be used instead as a reference for comparison with perturbed climates in the *piSST-pxk*, *a4SSTice* and *a4SSTice-4xCO2* experiments. We have updated Figure 1, Table 1 and the manuscript text to reflect this change. The rest of the protocol remains unchanged.

Editor's comments

Dear authors,

Thank you for submitting the revised version of your manuscript, "The Cloud Feedback Model Intercomparison Project (CFMIP) contribution to CMIP7" (egusphere-2026-398).

As also noted by the reviewers, this work is of great importance to the global climate modeling community. Coordinating a comprehensive and intricate community protocol like CFMIP4/CMIP7 is an exceptionally demanding endeavor, and I truly appreciate the substantial effort that the author team has devoted to developing and revising this manuscript. In general, the revised paper is in very good shape. To further enhance the clarity, depth, and utility of this protocol for the community, I would like to invite you to address the following remaining considerations before a final decision is made:

We thank the editor for his positive assessment and constructive feedback on the paper.

1. SST/SIC Forcing Datasets: One particular concern raised by both Reviewer 1 and Reviewer 3 pertains to the SST/SIC forcing data for the amip-piForcing experiments, including the sensitivity across different datasets and the recognized deficiencies in HadISST1.1, particularly over the Southern Ocean. For example, these deficiencies include uncertainties in clear-sky shortwave trends that may also be linked to trends in the tropics. These concerns motivated the adoption of the HadISST2 dataset for the CERESMIP experiment (Schmidt et al., 2023).

Given the Hadley Centre's plan to release HadISST2 in 2026 as the successor to HadISST1, it would be valuable to encourage modeling centers, where feasible, to conduct additional amip-piForcing experiments using HadISST2 if the dataset becomes available in time. I recognize that performing an additional set of experiments may be demanding for some groups, whereas other groups may be willing and able to undertake them. At a minimum, it would be beneficial to include a discussion of this issue and provide appropriate guidance for participating centers. It would also be informative to mention in this paper that HadISST2 has been proposed for use in the CERESMIP experiment, which can facilitate cross-MIP comparisons if the same model participates in both modeling experiments.

We agree it would be useful for modelling groups to run additional amip-piForcing variants with HadISST2, or other SST/sea-ice datasets as they become available – particularly since the choice of SSTs is known to affect the radiation budget. The editor is also correct that amip-piForcing simulations based on HadISST2 will be easier to compare with the amip simulations of CERESMIP.

We have therefore added a paragraph (from L231) discussing the upcoming HadISST2 SST dataset, with an encouragement for modelling groups to perform an additional third set of amip-piForcing simulations using this new dataset once available, and discussing the synergy with CERESMIP. Unfortunately, there is no confirmed date yet for the release

of HadISST2 as an operational product, so we are unable to include a timeline in the paper.

2. Discussion on the relationship of cloud processes to cloud feedbacks: Regarding Reviewer 2's comment on Section 2 (Page 4, L81), which emphasized the critical role of cloud processes themselves and how parameter perturbations contribute to cloud feedbacks and their overall spread across models, your response noted that "We have added some brief extra discussion of process studies, including the papers referenced here". However, an editorial check of the text reveals that these discussions and the specific references (Duffy et al., 2023; Gettelman et al., 2024) are not yet reflected in the revised manuscript. I suggest that these process discussions and citations be formally incorporated into the relevant part of Section 2 to strengthen the physical foundation of the protocol.

We thank the editor for spotting this and apologise for the oversight – it was our intention to add some discussion and references, but we somehow missed it. This is now rectified, with additional text and references (including the two mentioned above) in the first paragraph of the discussion of Q1 (from L75).

3. Additional Minor Refinements: Reviewer 3 also raised several additional comments (see the new review comments on revised version1), including the physical hypotheses underlying the experimental designs, the potential applications or alignment with recent field observations, and necessary clarifications regarding COSP diagnostics. Addressing these points will further improve the scientific contribution of the manuscript and support the robust implementation of the proposed experiments.

Please carefully address these points and submit your revised manuscript along with a point-by-point response. Once these technical and contextual updates are completed, the manuscript will undergo an editorial review for a final decision.

Our responses to Reviewer 3's suggestions are detailed in the point-by-point responses in the next few pages. In most cases we made changes to the text; in a few instances we found that the interesting points raised by the reviewer were somewhat beyond the scope of the paper, and our reasoning is explained in the responses.

We thank the editor once again and look forward to receiving his decision.

Reviewer 3

I commend the authors on a lucid and concise document. I also appreciate how difficult this can be for such a complex undertaking. Hence I offer my comments and invite them to consider them or not, as they wish, and as the editor sees necessary. I don't wish to impose requirements or additional burden on what is already a difficult enterprise.

We thank the reviewer for his positive assessment and constructive feedback on the manuscript.

The document nicely describes efforts to document how the latest efforts to fit global climate models to past observations influences differences in their behavior. In this regard, and given that there has been no real breakthrough in the way models are fit to data, isn't it important to acknowledge the need to document how they behave as compared to the models in past intercomparison projects, and which experiments contribute to that. If a model differed from what was inferred from the ensemble of models in CMIP5 would we be able to document why and how in the same way that we are trying to document how models in CMIP7 differ?

We agree this is an important point. We have added a paragraph in section 3 (from L167) to highlight that CFMIP and the CMIP DECK include a set of experiments that are essentially unchanged across CMIP generations (except for extending the time ranges). This facilitates comparisons across model generations to identify the reasons why the models differ. The assessment of Zelinka et al. (2020), cited here, is a good example of that.

I very much liked the emphasis on the science questions. While these helped understand how experiments were chosen to document the model behavior related to each question, I was wondering whether they are physically motivated. If so, highlighting the physical hypotheses that informed the experimental design would strengthen the scientific contribution.

The answer to the reviewer's point depends on what is understood by "physical motivation". Our thinking was that the discussion in the first paragraph after each question provides the physical basis. Furthermore, we note that our questions are deliberately broad in scope (to encompass the range of research activities across the CFMIP community) and therefore in most cases do not target any one particular physical mechanism. Thus for example, Q1 asks "What are the physical mechanisms…".

Under Q1, we have however added a specific mention to the "stability iris" hypothesis that motivates ongoing analyses of high-cloud amount feedback. We have also added a mention of the possible coupling between aerosol forcing and cloud feedback. For Q2 and Q3, we have included a few additional references that support the processes discussed.

Related to this, the manuscript acknowledges little how far physical understanding of state dependence has progressed in recent years. For example how the IWV influences the width of the window and hence the degree of water vapor masking (Bourdin et al., (2022)), how cloud masking of the surface window response depends on cloud temperature and coverage (Stevens and Kluft (2023)), and how the distribution of clouds in temperature space can be equifinal in CRE but lead to different radiative responses to surface warming (Kluft et al.,(2025)). These area all examples from my near field, but these ideas have also been highlighted in the work by Jeevanjee and collaborators.

We thank the reviewer for pointing this out, and for the literature references. We have added a sentence under Q2 (section 2, L100) to mention that state-dependence also affects non-cloud feedbacks and the degree to which they are masked by clouds, including the references cited above and a few more.

Table 1 is a nice overview, but it seems like some experiments don't relate to any question or application. Probably this is not true, but the manuscript gives this impression. Maybe it can be corrected by adding another category of purpose. There have been some very large and ambitious cloud and cloud circulation field studies over the past years, (AC)3, CAESAR, EUREC4A, ORCESTRA, OTREC. The manuscript gives the impression that the modeling community sees no value in these efforts. Certainly their relevance is not reflected in the paper, where I believe that they could guide the experimental design. Consider comparing ORCESTRA with GATE in AMIP, for instance.

This paragraph makes two separate points, which we address in turn.

1) Relating experiments to questions or applications: The fourth column of Table 1 lists the questions and applications that relate to each group of experiments. We are therefore unsure why the reviewer felt that that some experiments had no clear application. While we could add an additional column detailing the "purpose", this would essentially paraphrase the questions & applications already listed in the table. Perhaps the reviewer could clarify his thinking in case it was the presentation of the table that was confusing.

2) Cloud and circulation field campaigns: We acknowledge that field studies are essential for advancing understanding of cloud and cloud feedback processes, that they are a key interest of the CFMIP community, and that their importance could be better reflected in the paper. In fact, the CFMIP protocol is designed to be complementary to observations from a range of scales, from field campaigns (cfSites output) to global satellite observations (COSP output). This point is also related to another reviewer comment below on the relevance of process modelling. We have therefore added text in two places:

- Introduction: we have created a new (2nd) paragraph discussing the relevance of field studies as well as process modelling, including references to several recent field studies (L17);
- in section 4 (CFMIP data request), we now explicitly mention that the cfSites output requested for CFMIP is designed for comparison with field campaign

observations, and can be also used to derive large-scale forcings for process modelling (L326).

Aerosol was not mentioned, do they enter the protocol in any way? Given how strongly many models represent such effects, is there a strategy for documenting their impact on cloud responses. Or is there a reason (may be covered by other communities) that can be shared as to why not.

While aerosols are a stronger focus of AerChemMIP, RFMIP and RAMIP, we acknowledge that the CFMIP experiments and data request are relevant to the analysis of aerosol-cloud interactions, particularly where the physical processes overlap with those involved in cloud feedback and adjustment. We have added a paragraph to the conclusions discussing this (L353).

I didn't understand (§3.2.2) the use of HADSSTs. It made it sound like the AMIP SSTs are not being regularly updated. Given how emphasis the CMIP community has placed on updating the forcing datasets this seems hard to believe. Asking modeling centers to do extra work to assimilate other SST datasets just because the AMIP SSTs are not being updated misplaces the application of community persuasion. I do understand the value of looking at the sensitivity to small changes in SST patterns, but then a different experimental design would be more appropriate (climatological SST with HAD1SST and AMIP II SSTs)

It is indeed our understanding that AMIP SSTs are (for the time being) no longer regularly updated. This is because the NOAA OISSTv2.0 dataset, used to produce AMIP-II SSTs, was discontinued as of January 2023. The new operational version, 2.1, is fairly different from 2.0, and the CMIP community is yet to decide how AMIP-II will be extended.

We clarify that we have processed and made available HadISST1 in exactly the same format as AMIP-II SSTs, ready for use by modelling groups. Therefore we do not think this puts an additional burden on modellers. Furthermore, modelling centres are free to decide not to run the second *amip-piForcing* variant with HadISST boundary conditions.

As for the experimental design, this depends on whether the focus is on the cloud-radiative response for a particular time period (in which case we agree a climatological time-slice experiment is most appropriate), or the time variation. We find our proposal economical because it does not require an extra experiment, provides insight about the time-variation of feedback under different SST boundary conditions, and provides a useful test for candidate SST products to extend AMIP-II.

I appreciate that CMIP tends to focus on one type of model, and to the extent CFMIP is about designing experiments and output to help document the behavior of this type of model the experimental design works well. If however the purpose is to understand how clouds contribute to climate change (not just in CMIP models), a greater process focus would be beneficial. It would open the door to model diversity, from LES, to DNS, to kinematic cloud microphysical models, to storm-resolving models, to regional models, to observations.

We fully agree that process modelling is central to the science questions of CFMIP, and in fact the CFMIP community is involved in process modelling at a range of scales. See also our reply to the earlier reviewer comment on field studies. We now mention in the introduction that this paper is focused on global modelling, while acknowledging the relevance of process modelling (L17). We also highlight that the amip-p4k protocol is particularly suitable for high-resolution modelling, thanks to its low computational cost (L240). Additionally, we now also refer to process modelling in section 4 in the context of the data request (L326), where we highlight applications of CFMIP output for process models. Finally, we point out that the final paragraph of the paper highlights the informal MIPs supported by CFMIP, which include process-resolving model experiments (e.g. RCEMIP).

It is not clear to me where I would find the COSP diagnostics. Is this still the Swales et al paper? If so its use of ISCCP simulator output makes no physical sense. Clouds don't depend on height, and not really on pressure, but they do depend on temperature. If one wishes to distinguish between cloud temperature and surface temperature then this should be included in a dimension of the simulator, rather than pretending it is somehow captured by height. Not the least because passive sensors on satellites don't see clouds in this way, and the conversion to temperature space is straightforward for active sensors.

The full list of COSP diagnostics can be found in the CFMIP data request, which is referenced in the paper. As for the physical controls of cloudiness, while we agree that temperature could be used as a vertical coordinate rather than pressure, it is beyond the scope of the paper to discuss the representation of clouds in satellite simulators. For the purposes of CFMIP-related science, the key requirement is that cloud properties can be diagnosed consistently between models and observations; so any changes in satellite simulator output would require corresponding changes in the satellite retrievals.

Is there a strategy for dealing with the artificial treatment of clear-sky as 'cloud-cleared-sky' which is not observable. The latter arises because models don't treat the clear sky different in scenes where cloud fraction is less than one, which of course is wrong. One way to do this could be to output the radiation as histograms by cloud fraction, as that might get closer to what we observe, and it seems like that would be more desirable than contorting the observations to be wrong in the same way as the models.

While interesting, we also think that this point is beyond the scope of the present paper. We agree that clear-sky radiation, as defined in climate models, is not directly observable, making comparisons between observations and models challenging. We note that the CERES team implemented a new definition of clear-sky fluxes in their Energy Balanced and Filled (EBAF) product, from version 4.0 onwards (Loeb et al., 2020; 10.1175/JCLI-D-17-0208.1), designed to be consistent with the clear-sky fluxes as calculated in climate models, to facilitate comparison.

Maybe Li et al (2013) for the importance of accounting for slow ocean responses (line 168).

Thanks for the pointer, we have added this reference.

The use of the word “opportunity” rather than CFMIP data output “protocol” or “request” reads like a politically motivated change in language that a small group of people communicating in narrow circles has invented to help display their virtue. For most others I suspect it will just be confusing.

We sympathise with the reviewer’s dislike of the wording. However, diverging from the CMIP naming of the data request may confuse readers. We have noted at L176 that the word “opportunity” refers to the data request in CMIP7.