Reviewer Response - Round 1 - GMD
The Radiative Forcing Model Intercomparison Project (RFMIP2.0) for CMIP7

Reviewer #1

Summary

The authors present the experimental design and data request for the next iteration of RFMIP. The protocol is designed to answer four main motivating questions regarding the magnitude and components of present-day radiative forcing, how forcing components have evolved historically and will evolve into the future, how climate state influences forcing, and how land responses affect the diagnosed radiative forcing. Several of the proposed experiments follow from the CMIP6 iteration of RFMIP and will provide a consistent point of comparison with previous analyses; several new experiments are requested to address emerging questions; and several previous RFMIP experiments are no longer part of the official protocol (but are instead encouraged to continue informally). RFMIP serves as a crucial piece of CMIP, providing estimates of radiative forcing that are needed for a large number of scientific studies, including many done as part of other MIPs. The protocol described in this paper is well-motivated and should advance our ability to diagnose radiative forcings in models but also to gain insights into physical mechanisms underpinning them. Other than the fixed land surface temperature experiments (which are lower priority), the experimental design is relatively simple and the data request is fairly modest, making RFMIP a relatively low-burden MIP that yields high-value information. This review was jointly conducted by members of the Cloud Feedback Model Intercomparison Project (CFMIP) Scientific Steering Committee. We are highly supportive of RFMIP and recommend acceptance of this paper pending consideration of the comments below.

Signed,

Mark Zelinka, Paulo Ceppi, and Alejandro Bodas-Salcedo

We thank the CFMIP team for this thoughtful and helpful review. We've addressed their comments below in blue and within the updated manuscript. These updates have certainly strengthened the manuscript overall.

Major Comments

Question 1 can be understood as a special case of question 2. Only after reading Section 2.2 becomes clear that Q1 is specifically aiming at a more accurate characterisation of the present-day radiative forcing. It would be good to reword Q1 to highlight the emphasis on accuracy for the present-day forcing.

This is a good point. For Q1 we now say “What is the “precise” present-day radiative forcing …”

Regarding the framing of research question 4 on the impact of land temperature changes for forcing and feedback estimates, the paper explains that prescribed-SST simulations only approximate ERF, because land temperatures are free and they therefore induce a radiative response. This logic makes sense, but one could also argue that land warming is rapid (like other rapid adjustments) and therefore should be counted as part of the adjustments – in which case there is no need to correct for this land response when estimating ERF. Should the wording be a bit less categorical on this issue, for example at L122–124? One could for instance say that it’s not fully clear whether land warming should be treated as an adjustment, a feedback, or perhaps a mix of both. To be clear: the research to understand the impact of land temperature change on ERF is valid – it’s just that the framing could be adjusted.

We agree it would be good to be a bit less categorical here, given unresolved questions on this topic. We have softened the language. Line 126-130 now says “Though we do not advocate one particular definition of radiative forcing, since each has its own utility and limitations, in a strict sense, this is just an approximation of ERF. Land temperatures are able to respond in this standard fixed-SST approach and it is ambiguous to what extent the radiative effects associated with this are forcing adjustments or feedbacks.”

We left Q4 written as-is. We feel the phrasing “To what extent” already frames this as an appropriately open-ended question. And the above rewrite at the reviewers recommendation should further make this framing clear.

Minor Comments

● RFMIP2.0: use of decimal point seems redundant and not consistent with standard CMIP nomenclature (do you plan to run RFMIP2.x iterations?)

Given the increased focus on “operational” or routine model analysis, and the value RFMIP could contribute to these activities, we added the decimal to leave open the possibility of conducting smaller RFMIP iterations between CMIPs, as opportunities present themselves. If CMIP8 occurs, we would reset to RFMIP3.0. In part, this was inspired by DAMIP, who is also using the 2.0 nomenclature.

With this in mind, we have decided to stay with the RFMIP2.0 name, but welcome further thoughts from the reviewers.

● L47-48: Suggest also citing Zhou et al (2023) doi:10.1029/2022GL101700 here.

Yes, this is appropriate. We have added the reference

● L68: Should “impacts climate sensitivity” be “impacts our observationally-based estimates of climate sensitivity”?

No, our intent was to point to the value of RFMIP for estimating climate sensitivity more broadly than just observation-based analyses. It’s possible the citation we provided in that sentence was misleading. To alleviate this, we’ve added “e.g.” before the citation.

● L72. Here “effective climate sensitivity” is used, but “climate sensitivity” or “equilibrium climate sensitivity” is used on other parts of the manuscript. Please use consistent terminology. Also, the definition is missing, it would be worth defining it when equation (1) is introduced.

We have now changed all instances to “climate sensitivity” and added a definition near equation 1.

● L120–122: This way of estimating ERF is sometimes referred to as “Hansen forcing” in the literature (to distinguish it from regression-based “Gregory forcing), with a reference to Hansen et al. 1997 (10.1029/96JD03436). Suggest including that phrasing and reference here.

Good suggestion.We've added this to Line 124-125 along with the 1997 reference.

● L127: Given that substantial drift may be occurring in some models' piControl simulations, it may be better to enforce / recommend that the 30-year segment come from the portion of the piControl simulation near where various experiments (e.g., historical, abrupt-4xCO2) actually branch from it.

This is a good point. While internal to RFMIP this drift may not matter as long as the control SSTs are consistent with the SSTs in perturbed experiments. However drift effects could make the RFMIP experiments as a whole less relevant/consistent to other CMIP experiments (e.g. abrupt4xCO2) that RFMIP is often paired with. Given this, we have added an applicable recommendation to lines 131-133.

● L146: should "forcing boundaries" be "forcing boundary conditions" or just "forcings"?

We have changed it to forcing boundary conditions for consistency with the rest of the paragraph. This instance now occurs on line 153.

● L194: The DECK acronym has already been expanded in L76.

We thank the reviewers for catching this and have removed the expansion.

● L212: the Fast Track now seems to be referred to as "Assessment Fast Track" or AFT (cf Dunne et al. 2025 and CMIP7 website)

Correct. We have updated all instances of "Fast Track" to "Assessment Fast Track"

● L250-254: The comparison of state-dependence of ERF to the pattern effect for feedbacks is a little off or at least incomplete, because feedbacks also exhibit state-dependence independently of the pattern effect (e.g., 10.1038/s41561-020-00649-1, 10.1002/2015GL064240, 10.1029/2020GL089074).

This is true. We have added the references above and reworded the associated sentence (now Lines 265-267)

- L285: "an equal increase or decrease in CO_2 concentration". Please clarify that you are referring to a multiplicative increase/decrease.

We now say "an equal *multiplicative* increase or decrease" (now line 300)

- L287: strictly speaking, rather than non-linearity aren't you referring to a departure from logarithmic behaviour?

We agree non-linearity is not the best phrasing here. We now say "identify departures from logarithmic behavior" (now line 302)

- L317: Check reference format (B. Zhang et al); it was "Zhang et al" at L300

We have changed that instance of Zhang et al. to B. Zhang et al and we thank the reviewers for catching this.

- L422: "has become"

Revised (now L448).

- L442: It is discussed that fewer modelling groups have implemented the MODIS simulator. Given that MODIS is arguably a better instrument than the sensors that provide data to ISCCP, is less susceptible to misplacing low-level clouds at mid-levels, and has a record that is now more than 2 decades long, is it worth including more explicit encouragement to modellers to include MODIS simulator output?

Yes, we would like more modeling centers to use the MODIS simulator. We have revised the sentence (now L468-472) to say:

"We generally encourage modeling centers to implement the MODIS simulator in addition to the ISCCP simulator, given relevant cloud sensing capabilities of the MODIS satellite instruments and their long record length. But acknowledging that, traditionally, fewer modeling centers have implemented the MODIS simulator for CMIP, we stress that these MODIS diagnostics (and all other Tier B variable requests) are encouraged by strictly optional for participation in RFMIP2.0."

- Figure 1: More details are needed in the caption. How do these individual columns sum? Is anthro equal to the sum of ghg, aer, and lu? Is $1.4xCO_2$ a component of ghg? It is a little awkward that only the CO_2 component is labeled like this (with multiplication factor in front of it) – should it just be "co2" and then the explanation is given in the caption? The parenthetical statement “scaled from a 4xCO2 simulation to approximately 1.4xCO2 in 2014” is a little hard to parse and should be rephrased for clarity. Finally, use of colors seems superfluous for this figure.

These notes on our caption were very helpful. We have addressed each question and implemented each change mentioned. CO2 is indeed a component of ghg as now noted. And while ghg+aer+lu is a close approximation of “anthro” we do not expect it to be an exact match, in part due to non-linear interactions that occur within “anthro” in some models, and most importantly because “anthro” includes changes in ozone, which are not represented by the single-forcing-type experiments. This is also now explained in the caption of Figure 1.

We also added language in the main text (Line 239-242) to more clearly explain the decomposition and the handling of ozone.

- Tables 1,3,4: The caption states “minimum of 1 ensemble member” – but these are time-slice experiments; unless the spin-up is important, it would be simpler to output more years than more ensemble members. Perhaps instead state “minimum 30 years”?

We thank the reviewers for catching this. It was a typo and we only want to request one ensemble member, given that the 30yr timeslice setup itself is enough to minimize noise as intended. We have revised the Table 1, 3 and 4 captions accordingly.

- Experiment names: The current naming convention (judging from the Dunne et al paper) uses lowercase k: so e.g. piClim-p4K → piClim-p4k
- Table 3: piClim-4K-aer → piClim-p4k-aer (missing “p”).

We noticed that the Dunne et al. paper lower case “k” (e.g. in amip-p4k) is inconsistent with the Assessment Fast Track AirTable and past CMIP6 convention, which uses capital K for Kelvin. We checked with the lead author of that paper who confirmed it is indeed a typo and upper case K is the correct convention. Accordingly, we have kept the use of upper case K in the RFMIP experiment names. We have corrected the missing “p” in Table 3, however, and thank the reviewers for catching this.

Reviewer #2

In their manuscript “*The Radiative Forcing Model Intercomparison Project (RFMIP2.0) for CMIP7*”, the authors present the experimental protocol for the proposed RFMIP simulations as part of the CMIP7 project. The manuscript is clear and well written, and the experimental designs are overall well thought out and described, though I think the fixed LST simulations require additional guidance to ensure they provide interpretable, useful information for community analysis. The new RFMIP simulations proposed have excellent synergy with other MIPs, in particular with C4MIP, and demonstrate a refreshing appreciation for the role of the land surface in modulating global climate. I focus my review on the new RFMIP simulations proposed involving fixed LSTs, as that closely aligns with my area of expertise. I recommend the manuscript for acceptance following minor revision to address my comments below.

The fixed LST simulations described here will require "breaking" the surface energy balance; the authors should provide guidance to modelling centres on whether they should prioritize (a) energy conservation, (b) matching LST forcing, or (c) minimizing physically unrealistic turbulent fluxes.

A crude example of these fixed LST simulations potentially going sideways without appropriate guidance is as follows. Suppose the coupled land surface (freely varying T) has 10 units of energy to "get rid of" (to balance the surface energy budget... e.g. 5 coming in from shortwave radiation and 5 coming in from longwave radiation). In the control, freely varying state, perhaps the model puts 1 unit into ground heat storage, 5 into upwards longwave radiation, 2 into evaporation and 2 into sensible heat. Energy is conserved (10 "in", 10 "out"). For reference for the next case, lets also assume that there are only 2 units of water *available* to be evaporated at all.

Now suppose in the fixed LST simulation there is exactly the same energy coming in (probably there won't be, but for the sake of simplicity, let’s just assume that). 10 units of energy coming in. Now suppose the prescribed LST corresponds to an upwards longwave radiation of 11 units of energy. Uh oh. Where do we pull that energy from? Do we just make it up? That is an option. Alternatively, if modellers aren't told NOT to, maybe we try and get some extra energy in to conserve the 11 up from LST, by pulling heat up from the ground, or generating a negative sensible or latent heat flux. But those probably wouldn't be physically plausible with a weirdly hot land surface (which is what this example covers ... wouldn't get dew or downwards sensible heat *onto* something hotter than the air). But, that would conserve energy, if that is the goal. If it isn't the goal, what do we use for sensible heat, latent heat, and ground heat in these cases? Do we calculate them based on the prescribed LST of the last time step? This is what a model would do if it wasn't told not to. In that case, we'll have a really hot surface, so a big possible gradient in surface to air T, so a big possible sensible heat flux (upwards), so now we're *really* not conserving energy because let’s say now we have 4 units of sensible heat going up. Now we have 15 going up and only 10 coming down. Of course, we'd expect the atmosphere to warm up in response to this imbalance and maybe eventually equilibrate, but depending how this was set up it would be equilibrating to a simulation that has unexpected sources/sinks of energy at various times/places over land.

You could also see this causing physical inconsistencies in the latent/sensible heat fluxes, e.g. if the prescribed LST was large (hot land) which then heats the lower atmosphere which then leads to more evaporative demand and then the land runs out of water to evaporate. Then latent heat flux would go to zero and sensible heat flux would be come quite large (again possibly not conserving energy).

It is very common for land models, when doing energy conservation checks at the end of a simulated time step, to chuck any imbalance into the sensible heat flux term. Thus, if the modelling centres are not *explicitly told NOT to*, I think this experimental design is going to get some funky turbulent fluxes. In the example above, with the 11 units of energy up from the prescribed LST (corresponding to a hot land surface, which you'd intuitively expect to have upwards sensible heat flux), what many models would actually do is get to the end of the time step, check for energy conservation, notice it wasn't conserved, and use sensible heat as a tuning knob to conserve the energy – in this case, by making a downwards flux of sensible heat because that is what is required, even though it isn't physically realistic.

If the modelling centres *are* told not to adjust SH, we're going to get funky energy sources/sinks and lack of energy conservation. Including an additional variable (possibly variables plural) in these simulations that explicitly tracks energy imbalances could help those trying to make sense of the output understand what they're seeing. E.g. when/where is the land just spouting magic energy / sending it into the void?

I suggest the authors recommend that modelling centres

1. a) do/do not aim to conserve energy (authors should weigh pros/cons of each approach)
2. b) dump extra energy into sensible heat, or not (but provide a clear recommendation)
3. c) include a variable tracking energy imbalances (if conservation is not recommended)

We sincerely thank the reviewer for thinking with this level of detail about the framing and methodology of the RFMIP “fixed-land” experiments. Their clear explanation was very helpful for our understanding. We agree with the reviewer’s interpretation that fixing the land, while simultaneously still requiring the model to conserve energy, will lead to distinctive changes in turbulent fluxes as the land essentially becomes an infinite heat sink. However, we contend this is the intent of the experiments rather than a concern. The goal of these experiments is specifically to diagnose the effective radiative forcing and its adjustment components. Such changes in surface fluxes, and any effects of them, are a clear representation of the imbalances caused by the forcing agent and tell us precisely what effect the forcing is having on the atmosphere. For instance, the changes in sensible heat flux (and latent) are a requirement for restoring the tropospheric heat budget after introducing a forcing agent like CO2 (e.g. Andrews et al. 2010). Essentially, the turbulent flux changes are a signal of the adjustments that we are trying to isolate. We note that the same nuances regarding surface energy and susceptibility to turbulent flux adjustments occur over ocean in the standard, widely used fixed-SST experiment design.

Nevertheless, we agree it's important to be clearer about our goals and the reviewer outlines a good blueprint for doing so. We have rewritten the text (Lines 365 to 373) to make clear that we aim for the simulations to conserve energy, and despite likely leading to distinctive changes in turbulent fluxes, the model should be allowed to freely adjust its surface energy budget accordingly, as dictated by the model's own configurations (i.e. SH can adjust).

Andrews, T., Forster, P. M., Boucher, O., Bellouin, N., and Jones, A.: Precipitation, radiative forcing and global temperature change, Geophysical Research Letters, 37, https://doi.org/10.1029/2010gl043991, 2010.