Response to Referee and Community Comments

Response to Anonymous Referee #1 (RC1):

We thank the reviewer for the very helpful comments. Addressing these has improved the readability and clarity of the manuscript. Below we provide a point-by-point response to the comments. The original review is in bold text and our response is in regular text.

In this paper the authors seek to raise awareness of the contribution of external forcings, particularly historical, to climate model simulation uncertainties. Using the CESM large ensembles, including some new simulations using alternate CMIP forcings, they quantify the relative impacts of model changes and forcings in the CESM model framework. The newly presented CESM2-CMIP5 simulations are made publicly available as a resource for the wider research community to further investigate the impacts of forcings in the CESM large ensembles.

The paper is generally very well written, with clearly presented results supported by appropriate figures. The subject matter is a good fit for GMD and I recommend publications of this article subject to some minor changes.

Some particular issues to note are as follows:

1. **One has to be very careful when talking about the direction of fluxes – particularly when considering net fluxes and differences therein. In that regard the flux definitions in section 4 need to be made a bit more clearly with mention that net fluxes are considered positive down (or "downward").**

Thank you for raising this issue. We now clarify that fluxes are positive downward.

2. **Some of the ensemble naming is a little confusing. In particular it is a little confusing that "CESM2-LE" is used for the whole ensemble or for the 1st 50 members. It would be best to be a bit clearer here.**

We have taken the reviewer's suggestion and now use CESM2-LE to refer to the entire 100 member ensemble, CESM2-LEvbb to refer to the first 50 members which have variable biomass burning emissions for 1997-2015, and CESM2-LEsmbb to refer to the last 50 members which have smoothed biomass burning emissions.

3. **The 15 CESM2-CMIP5 simulations are introduced as being "newly presented" but it looks like DeRepentigny et al. (2022) has already presented results from 10 of these simulations (or are they different?). This point should be made somewhere. The fact that some of the simulations might have been used already elsewhere does not take away from the novelty of this paper.**

The reviewer is correct that earlier studies used subsets of these simulations to explore specific aspects of the climate system. These studies provided only limited details on the experimental design of these runs and the current manuscript provides additional details. We now clarify this within the introduction of the manuscript.

I include a pdf containing more detailed comments.

We have responded below to the comments that were embedded in the pdf.

Comments from pdf:

Line 83. A 10-member ensemble of CESM2-CMIP5 was already presented in DeRepentigny et al. (2022). So is it more accurate to say that you are extending that ensemble to 15 members rather than presenting new simulations?

The author is correct that a 10-member subset of these simulations have been discussed in DeRepentigny et al. (2022). However, the details on the experimental design of the runs were not covered in DeRepentigny et al. and the new manuscript presented here provides that documentation and so is meant to serve as a reference for these runs. To address the concern that these runs have been presented previously, we have removed the word “new” in several places in the introduction in reference to these runs. We also clarify on line 85 that “we present the experimental design of these CESM2-CMIP5 experiments”. We also note that subsets of these experiments have been used in some previous studies to explore some specific climate aspects.

Line 86. change to ", publicly available,"

Done

Line 115. Should "considered" be added here. You've not shown them to be not independent for the whole decade right? You just ignore that whole decade to be sure?

We have clarified that the 1920s are not “considered to be independent” here.

Line 130-131. So does this mean that hereafter "CESM2-LE" means only the 1st 50 members and "CESM2-LEsmbb" is the second 50 members? Or does your "CESM2-LE" include all 100?

If the former then that is a bit confusing given "CESM2-LE" is used for both and perhaps a new name is needed (CESM2-LEvbb?)? If the latter then that would seem a bit odd for your definition of BB influence in equation (3).

Thank you for pointing out that our description of the runs was not clear here. We now use CESM2-LE to refer to the entire 100-member CESM2 ensemble and use CESM2-LEvbb to refer to the first 50 members and CESM2-LEsmbb to refer to the second 50 members. Text and figures have been changed throughout the manuscript accordingly.

Line 138. You've already introduced "CESM2-CMIP5" above so why not use it here?

We have modified the text to now use it here.

Line 145. As in previous comment, you've already mentioned "CESM2-CMIP5" in the introduction.

We have removed this sentence since CESM2-CMIP5 is already introduced.

Line 150-151. Should this be the absolute imbalance? Or would a TOA of (say) -5 W/m^2 be ok?

We now clarify that this criteria is for the absolute imbalance.

Line 161. Is the "total" needed here? Or does it mean something special?

"total" does not signify anything special here and has been removed.

Line 171-173. Might this mean that, had you done the experiment the other way around, CESM1-CMIP6 might not have looked so different from CESM1-LE?
I.e., would the extra carbon mode make CESM2 more sensitive to biomass burning than CESM1?

This is an interesting point and we do believe that it is possible that the sensitivity to differences in biomass burning emissions could be different for CESM2 (using MAM4) than CESM1 (using MAM3). We now note this possibility within the text.

Line 187-189. Could you compare stratospheric AOD for both CESM1-LE & CESM2-LE to compare like-with-like?

Unfortunately, we do not have a consistent stratospheric AOD for CESM1-LE to compare with the CESM2-LE runs. As such, we have left the discussion and figure 2 as is.

Line 194-195. To make it clearer that the differences are between the models not the centuries, I would reword this to:
"...determine factors driving differences in 20th and 21st century warming between CESM1-LE and CESM2-LE."

Thanks. This has been reworded as suggested.

Line 204-205. One could argue that it is the 1997-2015 period that is "correct" given the variability is based upon satellite measurements. So are we missing some important physical processes in the before and after periods where the BB emissions are smoothed?

There are challenges with prescribing biomass burning emissions in climate simulations. Variable emissions as observed are in one sense more “correct” but they will not be consistent with the environmental state (meteorology, soil moisture, etc.) of each ensemble member and so in that sense are not “correct”. The lack of variable BB emissions throughout the simulations (for CESM1-LE, CESM2-LEsmbb, and the pre-1997 and post-2015 period in CESM2-LEvbb) is problematic given that previous studies have indicated the BB emission variability (and not just a smoothed mean) matters for some aspects of the climate. For the CESM2-LEvbb runs, the change in variability with the incorporation of BB emission satellite observations in CMIP6 is also problematic. Overall, it is a complicated topic. However, given that the BB story is only a minor aspect of the current manuscript, we don’t comment further on this here.

Line 216-217. Should you not exclude the 1920s having previously stated that they are likely not independent?

We now compute the differences relative to 1930-1950 although it makes only a very small difference.

Line 224-225. I guess this answers my earlier question about what is used for "CESM2-LE". This needs to be made clearer earlier. Perhaps "CESM2-LE" should be "CESM2-LEvbb" to contrast with "CESM2-LEsmbb"?

We have revised this as suggested and now use CESM2-LEvbb and CESM2-LEsmbb to contrast the first 50 and second 50 members of the CESM2-LE.

Line 228. This is a bit clunky. Although 2000-2020 is historical to us now, the period includes future scenario forcings - 14 years CMIP5 & 5 years for CMIP6.
I would change this to simply "Over the 2000-2020 period,"

We have replaced this with “For the 2000-2020 average,”

Line 239. What does this mean? Do you mean that net downward SW is positive? Best to include the word "downward" here and be clear that fluxes are downwards in general.

Thanks for pointing out that this was unclear. We now clarify that fluxes are positive downward.

Line 249. Need to reword this or add some punctuation

We have revised the wording and split into two sentences.

Line 250. Replace with "By"?

This has been revised.

Line 254. Why does the LWP change tend to 0 for CESM2-CMIP5 when both CESM1-LE & CESM2-LE end up at roughly the same location?

The LWP change tends to 0 for CESM2-CMIP5 by 2100 because of the compensation between model forcing and model structure and this is now explicitly mentioned in the text.

A comparison of the zonal-average LWP changes reveals that this is due to differences in the relative compensation of changing LWP in the tropics versus the polar regions for the different simulations. By 2100, LWP increases in polar regions in all simulations. Whereas tropical changes are small (in CESM1-LE) or somewhat negative (in CESM2-LE and CESM2-CMIP5). The tropical changes are most negative in CESM2-CMIP5 and this leads to a global average change that is minimal by 2100. Although we do not show this analysis within the paper, it is consistent with the comparison of zonal-average SW changes and cloud feedbacks shown in section 4.2.

Line 261. Add commas: ", in the 20th century,"

Done

Line 269-270. The deltas here look very small to be full depth-integrated change? Are you sure these aren't depth-averages?

The reviewer is correct that these are depth-averaged. Thank you for catching this. The text and figure caption have been revised.

Line 326. add comma

Done

Line 330-331. Presume you mean surface rather than whole column?

Yes. We now clarify that we mean surface here.

Line 337. This is odd wording. Suggest change to "From mid-21st century onwards..." or "From around mid-21st century..."

We have revised this to “From mid-21st century onwards”

Line 361. suggest add "the latter of" before "which" to make it clear you only mean forcing uncertainty rather than the combination

Modified as suggested.

Line 363. Although this might depend on how well the model internal variability matches the true natural internal variability of the system. If the ensemble is too confident then the spread might not include the observations from our "single reality".

We have now added the general caveat here that this depends on how well internal variability is simulated. Note that for CESM2, the model spread may actually be too large (at least for some climate properties) given that there is excessive power in ENSO.

Line 394. Add "scenario"?

Added as suggested.

Line 396. I would change to "used" because you've not really assess the SSP3-7.0 scenario per se, just the impact on CESM2.

Changed as suggested.

Line 400. This doesn't read well. Change "and" to either a new sentence or a semi-colon. Or add comma after "and".

We have split this into a new sentence.

Line 406. Suggest rewording to "..warming, the Arctic Amplification in CESM2-LE is considerably smaller than in CESM1-LE by 2100."

Modified as suggested.

Figure 1. Given that the focus here is on the historical forcings, it would be interesting to see exactly where the HIST-Scenario transitions lie because the x-axis scale is quite tight to pick out by hand. Perhaps a thin vertical line at the transition points (i.e., red at 2006, black at 2015) might be worth trialling?

Thanks for the suggestion. We have added a thin vertical line to the figure.

Figure 7. A minor point but I think it would be a little easier to see the latitudes if you had [0,30,60,90] as x-axis major ticks rather than [0,50]. Same for Figs 8, 9 & 10.

We have chosen to leave the figure x-axis ticks as in the original version.

Figure 7 caption. relative to 1920-1950?

This is now relative to 1930-1950 and this is now noted in the figure caption.

Figure 8. Although panels e-f are W/m^2, panels a-d have units of deg C. Are you sure this is SW TOA flux? I'm assuming this is just a typo on the axis label because the plots look sensible.

Thanks for noting this. It was indeed a typo and has been fixed.

Response to Anonymous Referee #2 (RC2):

We thank the reviewer for their helpful comments. We provide a point-by-point response to these below. In our response, the original comments are in bold text and our response is in non-bold text.

Review of "New model ensemble reveals how forcing uncertainty and model structure alter climate simulated across CMIP generations of the Community Earth System Model" by Holland et al. (2023)

Summary:
This paper describes a model ensemble for CESM2 with CMIP5 forcing (CESM2-CMIP5) and compares it with CESM1-LE and CESM2-LE. The comparison is able to separate the climate uncertainties caused by model structure and external forcings. They find a strong influence of historical aerosol forcing on the climate, and different forcing and model structure influences across the globe and regions. The paper is well-written, and the analysis is appropriate. It is an important effort to have these simulations for the community. I recommend publishing this paper in GMD after the comments below are addressed.

Specific comments:
1. The inclusion of CESM2-LEsmbb: This helps to demonstrate the impact of variable biomass burning in CMIP6 and is useful to explain the impact of forcing on SEP sea ice loss in the mid-21st century. However, the motivation to include this could be clearer. I suggest adding more discussion to reveal its effectiveness in interpreting the later results when introducing this experiment in L205.

We now clarify in the manuscript that one aspect of the total CMIP6 versus CMIP5 forcing influence quantified by Equation 2 is the inclusion of the variable biomass burning emissions from 1997-2015 in CMIP6 forcing, while CMIP5 forcing used smooth biomass burning emissions throughout. With the use of the CESM2-LEsmbb simulations, we can quantify the role that this plays in the CMIP6 versus CMIP5 emissions differences relative to the other CMIP5 versus CMIP6 forcing changes that are present.

2. Figure 6: Is this averaged over the whole ocean depth? I wonder if the model structure and forcing differences also affect the ocean responses in different depths. For example, have you had a chance to investigate the ocean temperature change from different vertical depths (e.g., upper, bottom oceans)? Is the response dominated by the upper ocean?

Yes. Figure 6 shows ocean temperature average over the whole ocean depth. We expect that there are certain depth ranges that dominate in the differences across model experiments. While this would be interesting to investigate, we believe that it is beyond the scope of the current work.

3. Cloud feedback calculation: I am wondering how the cloud feedback is calculated using the APRP method. Do you use the Gregory regression like processing the coupled abrupt-4xCO2 experiment? Or do you use the climatological TOA SW radiation anomaly divided by the global-mean and climatological surface air temperature anomaly? Please clarify it.

Described in Taylor et al. 2007, the Approximate Partial Radiative Perturbation (APRP) technique uses a simplified off-line radiative transfer model to calculate shortwave feedbacks. In this method, the simplified radiative transfer equations are used to estimate the local shortwave flux change attributable to clouds alone. This shortwave flux change due to clouds is then divided by the global mean surface temperature change to estimate the feedback. As detailed in Chalmers et al. (2022), APRP is especially useful to calculate the shortwave cloud feedbacks in polar regions because the APRP method incorporates the changing surface albedo into the calculation of shortwave flux change. As APRP is a standard technique used to calculate shortwave feedbacks for clouds (e.g., Zelinka et al. 2020, Chalmers et al. 2022) and the reference describing the method is already provided (Taylor et al. 2007), no revisions to the text were made.

Gregory regressions cannot be used to separate the change in shortwave flux attributable only to clouds. Therefore, we do not use Gregory regression to calculate shortwave cloud feedbacks.

References:

Chalmers, J., Kay, J. E., Middlemas, E. A., Maroon, E. A., and P. DiNezio (2022), Does disabling cloud radiative feedbacks change spatial patterns of surface greenhouse warming and cooling?, Journal of Climate, 1787–1807, DOI: 10.1175/JCLI-D-21-0391.1

Taylor, K. E., M. Crucifix, P. Braconnot, C. D. Hewitt, C. Doutriaux, A. J. Broccoli, J. F. B. Mitchell, and M. J. Webb, 2007: Estimating shortwave radiative forcing and response in climate models. J. Climate, 20, 2530–2543, https://doi.org/10.1175/JCLI4143.1.

Zelinka, M. D., Myers, T. A., McCoy, D. T., Po-Chedley, S., Caldwell, P. M., Ceppi, P., Klein, S. A., & Taylor, K. E. (2020). Causes of higher climate sensitivity in CMIP6 models. Geophysical Research Letters, 47, e2019GL085782. https://doi.org/10.1029/2019GL085782

4. L145-150: "Changing the forcing in CESM2 led to a slight radiative imbalance at the top of the model in the pre-industrial control, likely…": I wonder how large the imbalance is. Can you show the time series of piControl experiments with and without the tuning? How large is the radiative imbalance before and after the tuning? Please clarify it. Meanwhile, although the change of the tuning parameter seems to be small, its significant impact on global-mean radiative imbalance appears to suggest that it does matter. I wonder if the mean state climate, especially clouds, is altered significantly and the historical cloud simulations are modified. I encourage adding more discussions about this.

We now provide more information within the manuscript on the magnitude of the imbalance prior to and following the tuning. We also now quantify the difference in shortwave and longwave cloud forcing that results from the tuning, including providing additional figures within the supplemental material on the spatial differences in cloud radiative forcing.

5. Figure 4 and related figures with attribution plot: suggest adding the denotation of the uncertainty spread for 'Model' and 'Forcing' lines.

We have added information to the figure captions on what the range in the 'Model' and 'Forcing" on Figure 4 (and other attributions figures) represents.

6. Figure 5: units of LWP plots (c and d) are missing.

Thank you for notifying us of this typo. It has been fixed.

7. Figure 7: suggest adding the meaning of the black line in the caption for panel (e) and (f).

Done

8. L190: "... the eruptions have different relative forcing": What does the 'relative forcing' mean? Please consider rephrasing it.

We meant to say here that the relative forcing between specific eruptions differs in the CMIP6 and CMIP5 forcing. We have revised the text accordingly.

Response to Community Comment.

The community comment is shown in bold text below with our response in non-bold text.

"For historical and future simulations, this includes both natural (volcanic and solar) and anthropogenic (greenhouse gas, ozone, sulfate aerosol, and carbon aerosol) forcings. "

If the solar (sunspots, I am guessing since that is the mechanism y which solar shows uncertainty) is included as a forcing, why not tidal mechanisms? This has as large a forcing contribution to the highly reduced effective gravity of the ocean's thermocline, leading to climate perturbations such as El Nino and La Nina.

citations:

1. Pukite, P., Coyne, D., & Challou, D. ***Mathematical Geoenergy*** **(Vol. 241). John Wiley & Sons** ***(2019), Chapter 12 :Wave Energy*****.**

2. Lin, Jialin, and Taotao Qian. "Switch between el nino and la nina is caused by subsurface ocean waves likely driven by lunar tidal forcing." ***Scientific reports*** **9.1 (2019): 13106.**
Citation: https://doi.org/10.5194/gmd-2023-125-CC1

The commenter is correct that the variable solar forcing is associated with sunspots. Tidal forcing is not currently included in the standard forcings provided by CMIP5 or CMIP6 and so is not considered here. We do note though that the ocean model used within the CESM1 and CESM2 models includes a parameterization for tidal dissipation over rough topography (St. Laurent et al., 2002) as implemented by Jayne (2009).

References:

Jayne, S., 2009: The impact of abyssal mixing parameterizations in an ocean general circulation model. J. Phys. Oceanogr., 39, 17561775.

St. Laurent, L. C., H. L. Simmons, and S. R. Jayne, 2002: Estimates of tidally driven enhanced mixing in the deep ocean. Geophys. Res. Lett., 29, 2106, doi:10.1029/2002GL015633.