

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.
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?
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.
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.

5. Figure 4 and related figures with attribution plot: suggest adding the denotation of the uncertainty spread for 'Model' and 'Forcing' lines.
6. Figure 5: units of LWP plots (c and d) are missing.
7. Figure 7: suggest adding the meaning of the black line in the caption for panel (e) and (f).
8. L190: "... the eruptions have different relative forcing": What does the 'relative forcing' mean? Please consider rephrasing it.