

Geosci. Model Dev. Discuss., referee comment RC1  
<https://doi.org/10.5194/gmd-2021-298-RC1>, 2021  
© Author(s) 2021. This work is distributed under  
the Creative Commons Attribution 4.0 License.

## Comment on gmd-2021-298

Anonymous Referee #1

---

Referee comment on "Better calibration of cloud parameterizations and subgrid effects increases the fidelity of the E3SM Atmosphere Model version 1" by Po-Lun Ma et al., Geosci. Model Dev. Discuss., <https://doi.org/10.5194/gmd-2021-298-RC1>, 2021

---

This summarises (at some length) efforts to improve the calibration and evaluation of the atmospheric component of the E3SM coupled model. The procedure described seems extremely labor intensive and (frankly) somewhat arbitrary. Nonetheless, the results do show significant improvements and curiously, a reduction in the implied climate sensitivity (assessed via Cess-type perturbations). This is publishable with only minor revisions (as outlined below) and perhaps some condensing to reduce length and repetition.

I have two questions that might add to some of the discussion. What is the prospect for automating some of these tests, using ML/AI for instance to reduce the burden and increase the area of phase space tested? I don't have huge confidence that the current procedure will lead to true (local) minima in errors, but I'd like to see this discussed here.

Secondly, there is a preprint related to ECS in CESM2 (a related model), that has pointed out some odd (possibly erroneous) coding related to the ice-nucleation in that model (CAM6). Does this have any relevance here?  
<https://www.essoar.org/pdfjs/10.1002/essoar.10507790.1>

Minor points:

line 40. "...precise knowledge of ... ERF is not enough". This is a strawman argument. Who has ever said that it was?

line 84. The comparison to other ESMs is irrelevant. It the comparison to the constrained range from observations that matters (Sherwood et al, 2020).

line 114-115. Is there any evidence that the skill scores derived from a 5 day simulation are correlated to skill scores from a year or 10 year run? Presumably they are not being tested against the same observations?

line 120: "in hindsight"? is this referring to the 5-day simulations with EAMv1, or the previous one-at-a-time approach.

line 125. There is a big gap between 5 days and 10 years. Is there any assessment of how useful different lengths of simulation might be? For instance 1 year might be a good compromise?

line 128. use the actual times (10 years and 5 days) rather than 'short' or 'long' - relative measures are not very specific.

line 145. "perfect" is too much to ask. But the point about non-uniqueness is important.

line 160. Why? The authors just spent two pages saying why this was not a good approach!

line 224. "might"? --> "will"

line 245. Has the length of these simulations been mentioned?

line 385. How long are these simulations? (line 400 suggests 11 years, but is that just for simulation #5?) In any case, move this up in the text.

line 420/Table 6. Add observed values (where available) for comparison (i.e. from CERES, or CALIPSO).

line 438-449. please compare with Cesana et al (2021, doi:10.1029/2021GL094876). The implementation of a CALIPSO simulator should indeed be a high priority. Without a realistic target for LCF this tuning will inevitably be haphazard, but I think it likely that the EAMv1P is more realistic.

figure 4. The authors should add the EAMv1P-CERES map as well, so that the improved version can be compared directly with panel a. (also in Figures 5, 6, 9, 10, 11, and 12)

line 497. It is of course challenging, but I don't think the biggest challenge is the lack of observational data.

line 635. This will always be true - it is not a binary situation.

line 802. This is an odd argument. Who has ever claimed that ERF is sufficient to determine responses? It is precisely the opposite - the major uncertainty (since the Charney report!) has always been in the sensitivity.

line 818+. The comparison to the other models is fine, but the comparison should be with observationally constrained estimates - ie. Sherwood et al (2020), IPCC AR6 Chp. 7 etc.

line 1019. This has never been claimed.