Review of "Cloud properties and their projected changes in CMIP models with low/medium/high climate sensitivity"
By Bock and Lauer
egusphere-2023-1086

Summary
The authors compare climatological fields simulated by global climate models to those computed in observational datasets. The models are separated into high, medium, and low equilibrium climate sensitivity (ECS) categories, and it is found that for most fields examined, the high ECS models more closely resemble observations. Changes in these fields between the historical and future climate scenario are also examined, with the high ECS models showing largest changes in most location general. In general I was not impressed with this paper, for the reasons detailed below. I think the paper is flawed in its execution while also lacking a scientific motivation, and I therefore recommend rejection.

Major Comments

- The paper seems to lack any scientific question motivating the analysis or hypothesis that it is testing. Why are you evaluating these particular fields, and segregating the models by ECS? Is there a physical reason to expect the fidelity with which these fields match observations in the mean state to be tied to ECS? Do the authors believe that high ECS is more plausible than low ECS based on their results? What is the motivation for transitioning to examining how these fields change into the future? I didn't find any novel insights here that were not already well explained in the literature. In the end, I can't really understand what the point of the paper is, or why one would cite it.
- Most of the fields examined involve cloud properties (fractional area coverage, ice and liquid water path) or precipitation, but the evaluation is done without satellite simulators that ensure apples-to-apples comparisons of the geophysical fields (Bodas-Salcedo et al., 2011). It is well established within the community that one cannot simply compare a model cloud field to something retrieved from space, which has sampling biases, detection thresholds, scale differences, etc. Papers by Jen Kay, Greg Cesana, and others have made this point many times for several fields (G. Cesana & Chepfer, 2013; G. V. Cesana et al., 2021; J. E. Kay et al., 2012; Jennifer E. Kay et al., 2016, 2018). Even cloud radiative effect (clear- minus all-sky fluxes at the TOA) cannot be easily compared between models and observations because of differences in how clear-sky fluxes are provided in models vs observations (B. J. Sohn et al., 2010; B. J. Sohn & Bennartz, 2008; B.-J. Sohn et al., 2006). To facilitate more appropriate comparisons, adjusted clear-sky fluxes are now being provided by the CERES team (Loeb et al., 2020). For me, this decision to use raw model output to compare to satellite-retrieved fields is the most egregious flaw of the paper and I would need to see it remedied before I could recommend acceptance.
- I found it very disconcerting that the authors did not ensure a common time period for their model-observation comparisons. Why are the climatologies from the various observational products different from each other and from the models (1985-2004)?
- The Observations section was literally 3 sentences, none of which actually explained the datasets, their version/collection, nominal resolution, what instrument (on which satellite) is

measuring each geophysical quantity, etc. This is unacceptable for a scientific manuscript in which models are being evaluated against observations. The recurring cloud product with the acronym ESACCI is not even defined anywhere.

- In stark contrast to the 3-sentence Observations section, Section 2.3 reads like an advertisement for the ESMValTool. Most of this information regarding the software you used to perform your analysis is meant for the *Code and data availability* section.
- The changes in cloud properties are computed by differencing the future scenario with the historical scenario. While this will provide the total change in clouds, those changes will be due to an ambiguous mix of causes: responses to warming, decreases in aerosol loading, and adjustments from changes in other forcing agents. High ECS models typically also have large aerosol-cloud interactions (Kiehl, 2007; Wang et al., 2021), so a portion of their change between historical and future climates will be due to a recovery from being strongly affected by aerosols in the historical period, and will not be purely attributable to cloud feedback processes.

Specific Comments

- Author list: both authors' names are in reverse order
- L7 and throughout: "both, cloud physical" the comma after both is not needed; this typo recurs throughout the paper (e.g., L26, 140)
- L99: what simulations are being used here? Also, it should be caveated that the change in cloud radiative effect is not the same as the cloud feedback owing to changes in clear-sky fluxes that are not related to clouds (Soden et al., 2004)
- L111: it doesn't matter which direction one is going; delete "when going from south to north"
- L111-113: these statements are made without providing any evidence of the role of changing cloud phase; suggest either deleting, citing the appropriate literature, or providing evidence.
- Figure 1: Given that ECS is strongly dependent on cloud feedback, it seems odd to plot cloud feedback on the y-axis, which is typically thought of as the dependent variable.
- L209: "clouds are warming" should be re-stated
- L226 vs L227: "largest positive bias"…."too strong net cloud radiative effect" – I'm confused about what these mean. The net CRE is negative, so if it is "too strong" I'd expect that to mean that the negative magnitude is too large, but this would not be a positive bias. Please restate.
- Figure 10 and elsewhere: I'm not sure what is meant be "relative change". How is this computed?
- Figure 11: if liquid water path is denoted as lwp rather than clwvi, it seems that ice water path should be denoted as iwp rather than clivi. Should one care about IWP over the stratocumulus regimes?
- L383-384: In this sentence, every possible regime on the planet is listed; is this really informative or helpful? If you quantified more rigorously the regimes that are strong contributors to inter-model spread in cloud feedback or ECS, you would find that not every location on the planet contributes equally.

References

Bodas-Salcedo, A., Webb, M. J., Bony, S., Chepfer, H., Dufresne, J. L., Klein, S. A., et al. (2011). COSP Satellite simulation software for model assessment. *Bulletin of the American Meteorological Society*, *92*(8), 1023–1043. https://doi.org/10.1175/2011bams2856.1

Cesana, G., & Chepfer, H. (2013). Evaluation of the cloud thermodynamic phase in a climate model using CALIPSO-GOCCP. *Journal of Geophysical Research: Atmospheres*, *118*(14), 7922–7937. https://doi.org/10.1002/jgrd.50376

Cesana, G. V., Ackerman, A. S., Fridlind, A. M., Silber, I., & Kelley, M. (2021). Snow Reconciles Observed and Simulated Phase Partitioning and Increases Cloud Feedback. *Geophysical Research Letters*, *48*(20), e2021GL094876. https://doi.org/10.1029/2021GL094876

Kay, J. E., Hillman, B. R., Klein, S. A., Zhang, Y., Medeiros, B., Pincus, R., et al. (2012). Exposing Global Cloud Biases in the Community Atmosphere Model (CAM) Using Satellite Observations and Their Corresponding Instrument Simulators. *Journal of Climate*, *25*(15), 5190–5207. https://doi.org/10.1175/JCLI-D-11-00469.1

Kay, Jennifer E., L'Ecuyer, T., Chepfer, H., Loeb, N., Morrison, A., & Cesana, G. (2016). Recent Advances in Arctic Cloud and Climate Research. *Current Climate Change Reports*, *2*(4), 159–169. https://doi.org/10.1007/s40641-016-0051-9

Kay, Jennifer E., L'Ecuyer, T., Pendergrass, A., Chepfer, H., Guzman, R., & Yettella, V. (2018). Scale-Aware and Definition-Aware Evaluation of Modeled Near-Surface Precipitation Frequency Using CloudSat Observations. *Journal of Geophysical Research: Atmospheres*, *123*(8), 4294–4309. https://doi.org/10.1002/2017JD028213

Kiehl, J. T. (2007). Twentieth century climate model response and climate sensitivity. *Geophys. Res. Lett.*, *34*. https://doi.org/10.1029/2007GL031383

Loeb, N. G., Rose, F. G., Kato, S., Rutan, D. A., Su, W., Wang, H., et al. (2020). Toward a Consistent Definition between Satellite and Model Clear-Sky Radiative Fluxes. *Journal of Climate*, *33*(1), 61–75. https://doi.org/10.1175/JCLI-D-19-0381.1

Soden, B. J., Broccoli, A. J., & Hemler, R. S. (2004). On the use of cloud forcing to estimate cloud feedback. *J. Climate*, *17*, 3661–3665. https://doi.org/10.1175/1520-0442(2004)0172.0.CO;2

Sohn, B. J., & Bennartz, R. (2008). Contribution of water vapor to observational estimates of longwave cloud radiative forcing. *J. Geophys. Res.*, *113*. https://doi.org/10.1029/2008JD010053

Sohn, B. J., Nakajima, T., Satoh, M., & Jang, H. S. (2010). Impact of different definitions of clear-sky flux on the determination of longwave cloud radiative forcing: NICAM simulation results. *Atmos. Chem. Phys.*, *10*(23), 11641–11646. https://doi.org/10.5194/acp-10-11641-2010

Sohn, B.-J., Schmetz, J., Stuhlmann, R., & Lee, J.-Y. (2006). Dry Bias in Satellite-Derived Clear-Sky Water Vapor and Its Contribution to Longwave Cloud Radiative Forcing. *Journal of Climate*, *19*(21), 5570–5580. https://doi.org/10.1175/JCLI3948.1

Wang, C., Soden, B. J., Yang, W., & Vecchi, G. A. (2021). Compensation Between Cloud Feedback and Aerosol-Cloud Interaction in CMIP6 Models. *Geophysical Research Letters*, *48*(4), e2020GL091024. https://doi.org/10.1029/2020GL091024