Review of “Cloud properties and their projected changes in CMIP models with low/medium/high climate sensitivity” by L. Bock and A. Lauer.

In this study, the authors classify outputs of 51 CMIP5 and CMIP6 models into low, medium and high ECS groups and then compare them with observations. They further look at the change in cloud properties between historical and 4xCO2 simulations weighted by global mean surface warming. They find that the models from the high ECS group better represent the observed climatology of cloud-related variables and have different sensitivities to warming than the low- and medium-ECS groups.

The topic of this paper aligns well with the scope of the journal. While I recognize the value of investigating the response of clouds to warming and the amount of work it takes in terms of data processing, I find that the study suffers from two major flaws: a non-consistent direct comparison of model cloud-related outputs with observations and the use of coupled historical simulations in the model evaluation. Also, it seems that the authors have already evaluated these models in a separate paper, so I see little value in doing this again. However, I find the analysis of the cloud response to climate change very interesting. More details are given below.

Main comments:
My biggest concern is the comparison of cloud-related fields with observations, which doesn’t account for observational uncertainties and inherent limitations of the satellite instruments. The LWP products suffer from large uncertainties (sometimes several times greater than the observed value itself, Lebsock and Su, 2014; Elsaesser et al., 2017) and cannot be used to assess models on a global scale. IWP products seem to more reliable but there is still the question of whether precipitation is accounted for or not (e.g., Li et al., 2014). The cloud fraction also cannot be compared directly to observations because of the instrument limitations and the difference in cloud definitions between models and observations. I’m attaching a figure showing the impact of using ISCCP (basically AVHRR), MODIS and CALIPSO simulators on the original output of the model for 3 CMIP6 models. The differences are very large, region dependent and model dependent…

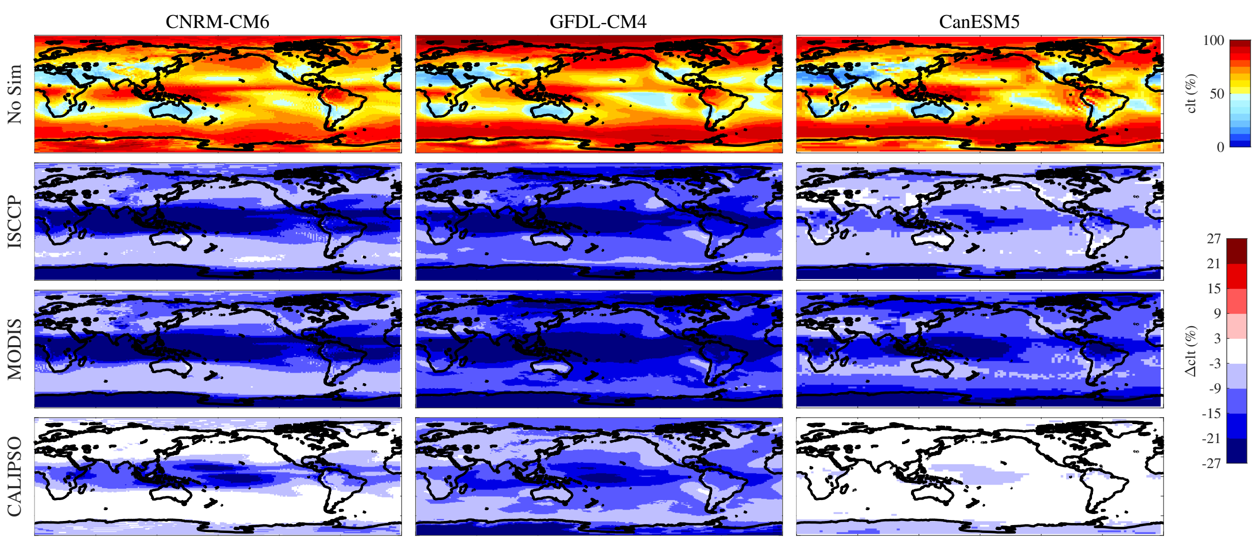

Figure 1: Effect of ISCCP, MODIS and CALIPSO simulators on total cloud cover in three CMIP6 models. Total cloud cover (‘clt’) as simulated by CNRM-CM6, GFDL-CM4 and CanESM5 CMIP6 models (first row) and the difference between the original total cloud cover and that simulated by ISCCP (second row), MODIS (third row) and CALIPSO (fourth row) simulators.

The second main concern is the comparison of historical simulations with present-day observations that have not the same surface forcings. The SST pattern and magnitude, which have strong impact on all the variables that are studied here, are not well reproduced by the coupled models as shown in the literature (e.g., Seager et al., 2019). Therefore, it is not a fair comparison. Instead, the authors should use AMIP type simulations to assess the models.

Another main comment, which could be easily fixed, is the conclusion. Except for the last paragraph, which is very insightful, the conclusion is far too long (2 pages) and does not summarize the results but rather re-state them without any apparent structure.

Minor comments:
Almost no information about the observations used is given and including potential uncertainties, which are raised here and there in the manuscript but without being formally quantified. As is, it looks like the authors have very little knowledge about the observations they're using.

I couldn't find a clear definition of how the feedbacks are computed.

The introduction is not doing the best job at motivating the ECS discrimination. If the idea is that larger cloud feedback could be related to mean state cloud properties, then I would classify the models by the global mean feedback rather than ECS, because the ECS is affected by other feedbacks than those from clouds. The way it is presented in the paper is even slightly backward in my opinion.

I question the usefulness of having 2 to 3 versions of a model with the same atmospheric component especially when it comes to evaluating atmospheric quantities. They have disproportional impact on the mean. This question is not specific to that study though.

L129-130: then why is there no distinction between CMIP5 and CMIP6 models?

L136-139: I'm not sure what this means. There can be ice clouds at all latitudes in the high levels. This latitude loosely corresponds to where clouds can be mixed-phase cloud almost year-round. These clouds show different feedbacks from the warm clouds.

L158: simulated is written two times in a row and I don't think that Z20 is saying or showing such a conclusion in their study.

Fig. 3: Aside from the non-consistent model-to-observations comparison, the spread and SD appear to be very similar between the group, so the differences are not significant and there is no clear systematic behavior among the groups.

L198-199: I don't understand this sentence.

L243: not the number on the figure, so I suppose the authors decided to switch the default dataset to CERES during the first round of review, but failed to revise the text.

I don't see the added value of Fig. 10 and 11 compared to Fig. 9. I think that Fig. 9 and the analysis associated to it is the best part of this manuscript and should be the focus of this study.

Each main Sc decks is singled out yet no there is no motivation for doing this, I'm also not a fan of picking fixed regions to study cloud response to warming since these decks can evolve in terms of location.

The beginning of the conclusion is confusing. The authors argue that the increase of ECS between CMIP generation motivated them to investigate the response of clouds to climate change, yet most of the study is based on present-day climate evaluation and they do not segregate between CMIP5 and CMIP6.

Line 382: Z20 do not say this, instead they argue that this is a possibility that should be investigated.

References:

Elsaesser, G.S., C.W. O'Dell, M.D. Lebsock, R. Bennartz, and T.J. Greenwald, 2017: The Multi-Sensor Advanced Climatology of Liquid Water Path (MAC-LWP). J. Climate, 30, no. 24, 10193-10210, doi:10.1175/JCLI-D-16-0902.1.

Lebsock, M., and Su, H. (2014), Application of active spaceborne remote sensing for understanding biases between passive cloud water path retrievals, J. Geophys. Res. Atmos., 119, 8962–8979, doi:10.1002/2014JD021568.

Li, J.-L. F., Forbes, R. M., Waliser, D. E., Stephens, G., and Lee, S. (2014), Characterizing the radiative impacts of precipitating snow in the ECMWF Integrated Forecast System global model, J. Geophys. Res. Atmos., 119, 9626–9637, doi:10.1002/2014JD021450.