Review of “**Reconciling Satellite-Model Discrepancies in Aerosol-Cloud Interactions Using Near-LES Simulations of Marine Boundary Layer Clouds**” by Shaoyue Qiu, Xue Zheng, Peng Wu, Hsiang-He Lee, and Xiaoli Zhou

This manuscript identified model bias on LWP responses to aerosol perturbations and potential causes behind such bias using near-LES simulations and multiple observations over the Eastern North Atlantic region. By comparing the modelled LWP susceptibility with satellite observations, they found that modelled LWP susceptibility from non-precipitating, thick clouds have the largest discrepancy compared to the observations, while the LWP susceptibilities from precipitating and non-precipitating thin clouds show relatively good agreements with observations. It is suggested that the model overestimates precipitation for thick clouds including excessive autoconversion and accretion, and underestimates entrainment and evaporation, which are the main reasons for the LWP susceptibility discrepancy in these non-precipitating thick clouds. They also found that the modelled cloud susceptibilities are sensitive to cloud top humidity, and the bias of cloud top humidity in the model can be another reason for the LWP susceptibility discrepancy.

The findings in this manuscript are insightful and important for improving representation of aerosol-cloud interactions in the models. The topic and research questions are also relevant within the scope of ACP. However, I have several major comments outlined below for the improvement of this manuscript, and I recommend resubmission after the following comments are addressed.

Recommendation: major revisions

Major comments:

- I am concerned about the ability of the model to simulate LWP for the selected cases. In Figure 5, the model simulates non-precipitating, thick clouds with high LWP much more frequently than the Meteosat observed. These non-precipitating, thick clouds are key to the later-on analysis and conclusions. Comparison of LWP between model and observation is only for two cases, and Figure S2-S4 only provide a qualitative comparison of cloud fields. I suggest a more quantitative model-observation comparison for the selected cases, and a more detailed description and explanation on the LWP bias (currently there is only one sentence at Line 422 stating the potential reason of lack precipitating scavenging feedback on aerosol and N_d) and how this bias affects your conclusions. Although constant aerosol number concentrations are used for simulations, it will be helpful to have the N_d comparison as well.

- The “Data and methodology” section needs more details. For observational data, what are the specific products or variables used from satellite? What are the uncertainties of your observations and how good are they? How did you calculate N_d from Meteosat, what is the assumptions and uncertainties of the selected method on your cases?

For WRF model, how are the key warm cloud processes treated in your model, what are the parameterizations and what are the limitations of these treatments for your cases? What is the limitation of using a constant total aerosol number concentration throughout the domain for your model-observation comparison on LWP susceptibility? What is the default value you selected for aerosol number concentrations for your cases and are they the same for all cases? How did you quantify the N_d-LWP relationships driven by internal cloud processes and by cloud base updraft speed?

- Naming of model simulations are unclear and sometimes confusing throughout the manuscript. Currently they are described with “polluted” and “clean” in comparison. This can be misleading when you switch to another set (e.g., N=500 can be “clean” compared to N=1000 but can be “polluted” compared to N=100). In addition, “clean” is also used for describing observations (Line 497) and there is also a description of “ultra-clean” (Line 560) for the N=100 simulation. I suggest a consistent name for each model configuration in the manuscript for clarity.

- How do different synoptic regimes affect the LWP susceptibility? You mentioned to investigate the variation of ACI across different synoptic conditions in the Introduction (Lines 117-120) and therefore chose these 11 cases, however little results and analysis are shown in this manuscript on this question.

- Many captions in this manuscript are not complete and refer to captions in another figure. I suggest to include full captions for all the figures and be clear about the data used in the figure.

- In Section 3.3.1 Precipitation Efficiency, there are many comparisons between model and ground-based observations for cloud with different R_e and optical depth. However, the current Figures 7-10 are for observations, N=100, N=500, N=1000 and each has 9 subplots categorized by R_e and optical depth, making the whole section sometimes hard to follow. It might be helpful to reorganize these figures and perhaps paragraphs as well, so that observation and all model results are in the same figure for comparison. For example, Figure 7 can just contain clouds with optical depth less than 10 and the column now becomes observation, N=100, N=500, and N=1000. Or separate the figures by non-precipitating, drizzle and rain.

Minor comments:

- Line 1: I don’t think “reconciling” is accurate for the title of this manuscript. I think key processes and reasons behind the inconsistent LWP susceptibility are identified in this manuscript, but this issue is not resolved here and requires model improvement.
- Lines 18-19: “largely due to” – I don’t think incorrect LWP responses to aerosol perturbations is the reason but a main issue. The reasons can be poor representation of aerosol and cloud processes.
- Line 25: “a modest LWP decrease” to an increase in N_d.

- Line 26: "In contrast" to? It feels coming from nowhere. If you would like to suggest that non-precipitating thin clouds have consistent LWP susceptibilities from model and observation, but not for non-precipitating thick clouds, then you need to state this clearly.
- Line 108: please define the abbreviation of "MBL".
- Lines 128-133: What are the specific cloud retrievals and what are the uncertainties of each cloud retrieval? In addition, you have the method of calculating N_d from satellite mentioned at Line 386-392, but I think it will be better to move to this section. It is also useful to include version numbers of satellite product here and in the Data availability section.
- Line 138: How was the satellite retrieval smoothed to 25-km resolution?
- Line 155: "0000 UTC"
- Lines 163-169: ERA5 data is not observational data but reanalysis data, therefore I don't think this should be described here under the observational data subsection. It can be put in a separate subsection, or you can change the name of this subsection to something like "Datasets" and separate into satellite data, ground-based data and reanalysis data.
- Lines 180-182: What are the spatial resolution of the other two nested domains?
- Line 186: How often is the lateral boundary condition updated?
- Lines 189-191: How are boundary layer and clouds treated in the innermost domain?
- Figure 1: How does Meteosat retrieve cloud coverage and is the modelled cloud cover comparable to the Meteosat-retrieved cloud coverage? How is cloud top height defined in model output and how does Meteosat retrieve cloud top height? I suggest adding time series of N_d here. In addition, how does N=500 simulation look like?
- Lines 291-292: I don't think the cloud coverage from N=100 simulation closely matches the observed cloud coverage, but underestimates the cloud cover. It will be helpful to add some numbers here as well, rather than just quantitative descriptions.
- Lines 292-294: Can you suggest the reasons behind the model failed to simulate the dissipation of clouds? And how may this bias affect the modelled LWP susceptibility?
- Figure 3: Please use a full caption here rather than referring to another figure's caption. Similar to the comments for Figure 1, I suggest adding time series of N_d here as well.
- Figure 4: Please use a full caption here.
- Figure 5: "WRF simulations" are these from all polluted versus clean simulations or just one of the sets? Are R_e on these plots from the model or from satellite? Please make sure the axes are same for the model and observation plots. Currently they are different and make it difficult to compare with.
- Line 382: How does the Meteosat LWP susceptibility calculated?
- Line 386: What does it mean by "to be consistent with satellite observations"?
- Line 395: I think it will be useful to add a sentence here on how you define different types of clouds: precipitating versus non-precipitating, thick versus thin.
- Lines 407-410: Your satellite observations for precipitating clouds are different from your simulations and previous study with long-term data. Can you suggest why? Is this because of the limitations of satellite data? Does this affect your model-satellite comparison for other clouds?

- Line 429-430: I don't think Figure 5 show that the model results agree with Meteosat observations for an increase in LWP in precipitating clouds (Meteosat suggest a decrease).
- Line 434-435: If the modelled LWP response is showing large discrepancy compared to observations, this is not indicating the robustness of the results. Please explain in detail on the reasons why you suggest that the model results are robust.
- Figure 6: It is confusing here that the r_e dashed lines across different r_e contour colours in (a) and (c). Please be clear about how each effective radius is calculated or derived in (a), (c) and the dashed line.
- Lines 493-495: Frequencies from satellite data only sum to 90.6%, what and where are the rest 9.4%? In addition, can you explain more on why the selected cases are representative just based on the frequencies?
- Line 497 and others: what does "clean condition" mean here? You use "clean" to describe both simulations and R_e condition in your figures in this section, which is confusing during reading.
- Lines 506-507: "likely due to mixing and evaporation" – can you be more specific on this?
- Figures 8-10: Please use full captions for these figures.
- Lines 558-559: I can see that DSD is compared by using percentages of R_e categorizes, but it may be helpful and clearer to compare full DSD from different model simulations and observations for clouds with different optical depths.
- Lines 591-599: The description of DSD in the model is better to be put in the Data and Methodology section along with the descriptions of other treatments of warm cloud processes.
- Lines 636-638: The cloud tops are defined differently in ARM observations and in the model. Since you have the model radar simulator, why not using the same definition here based on the radar reflectivity profile for observed and modelled cloud tops?
- Lines 644-648: I commend the authors on considering the spatial representation issue and it will be helpful to describe how the temporal representation issue is treated, e.g., what are the model output time for comparing cloud top RH with the sounding observations?
- Figure 12: What is the shaded area for?
- Line 656: "in the simulations"- are these for all simulations with all aerosol number concentration or specific ones? Does the dependence of these cloud susceptibilities on cloud top relative humidity change when using different sets of simulations (e.g., between N=1000 vs. N=100 and N=1000 vs. N=500)?
- Figure 13: It will be helpful to have vertical lines where buoyancy flux difference equals to 0 as well on the plot. Similar to the comment on Figure 12, what is the shaded area for?
- Line 691: please explain the N_d-LWP here in detail.
- Figure 14: Are these from simulations with all different aerosol concentrations?
- Figure 15: Please use a full caption.
- Lines 745-754: I think several references are missing here in this first paragraph when mentioning the findings from previous studies.
- Lines 762-763: I suggest adding the LWP bias here rather than using "generally match".

- Lines 778: Are there any other potential reasons for the LWP bias and what's the reason that you suggest the lack of precipitation scavenging feedback on aerosols is likely the cause here?