Responses to reviewer comments

In this response letter, we provide a point-by-point response to each comment. The original comments are in *blue italic font*, and our response are in black font. Changes made in the manuscript are listed in quotes with line numbers of the tracked changes version.

Reviwer #2:

General comments：

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 Nd) and how this bias affects your conclusions. Although constant aerosol number concentrations are used for simulations, it will be helpful to have the Nd comparison as well.

Thank you for the constructive comment. In Figure R1, we include a quantitative comparison between model and observed LWP and Nd. As seen in Figure R1a, the simulated LWP in the N=100 simulation agrees well with the Meteosat observation, with a mean value about 10% lower than Meteosat. However, since most of the thick clouds with LWP greater than 75 g/cm2 exhibit a positive LWP susceptibility (Figure 5a), the LWP in the N=500 and N=1000 simulations increases relative to that in the N=100 simulation.

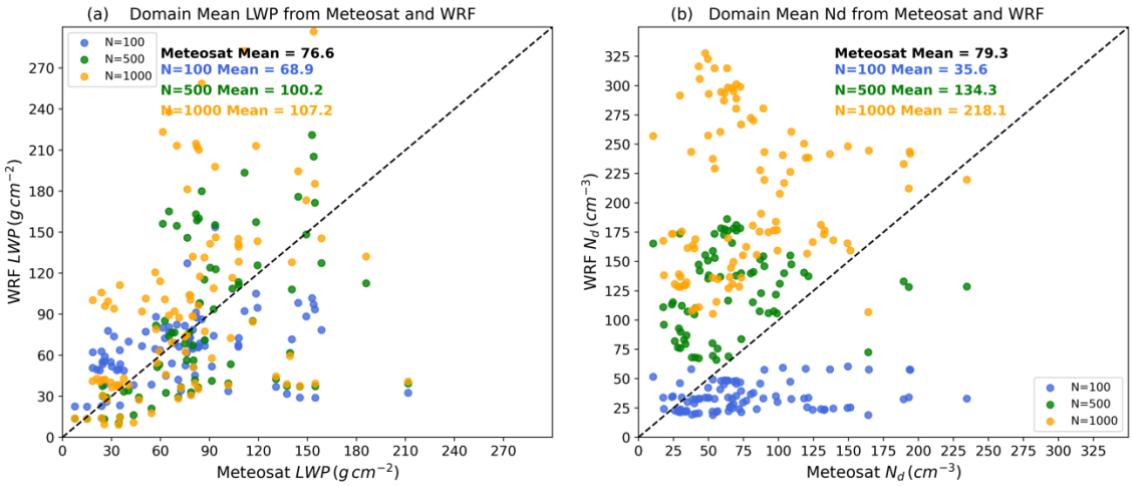

Figure R1 Scatter plot of domain-averaged cloud properties from Meteosat observations and WRF model simulation. Different colors represent different simulations (N=100, blue, N=500, green, N=1000, orange). (a) cloud liquid water path (LWP), (b) cloud droplet number concentration (N_d).

The overestimations of N_d are due to the overestimated prescribed aerosol concentration in model setting combined with the lack of precipitating scavenging effect on aerosols in these simulations. Based on field campaign measurement, the mean total aerosol concentration (N_a) and N_d in the ENA region in summer are ~400 cm^{-3} and 80 cm^{-3} respectively (Zhang et al., 2021; Wang et al., 2021; Wang et al., 2022; Zheng et al., 2024). Since the purpose of this study is to quantify aerosol-cloud interactions and cloud susceptibility in model simulations, the prescribed aerosol concentrations are designed as N=100, N=500, and N=1000 to sufficient variation in N_a. The simulated Nd in these simulations are lower and higher than Meteosat retrieved N_d (Figure R1b). We have added the model-satellite comparison of LWP to the supplement, and added the following text to the manuscript:

"The overall overestimation of N_d likely arises from the prescribed aerosol concentration used in the model configuration, combined with the absence of precipitation scavenging. For reference, the mean aerosol concentration over the ENA region during summer is approximately 400 cm^{-3} (e.g., Zhang et al., 2021; Wang et al., 2021; Wang et al., 2022; Zheng et al., 2024). The model's overestimation of LWP may stem from its excessively positive LWP susceptibility in thick clouds. As shown in Figure S9, simulated LWP in the N=100 simulation agrees reasonably well with the Meteosat retrieval, with a mean value about 10% lower than observed. However, in the N=500 and N=1000 simulations, the strong positive LWP susceptibility leads to increases in LWP for clouds with LWP> 75 gm^{-2}, resulting in mean values 30% and 40% higher than Meteosat retrievals, respectively. (Lines 556-565)"

- 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 Nd 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 Nd-LWP relationships driven by internal cloud processes and by cloud base updraft speed?

Thanks for the questions and detailed suggestions. To address this comment, we have added the following text to the manuscript:

"The SEVIRI Meteosat cloud retrieval products are pixel-level cloud retrievals produced by NASA LaRC SatCORPS group, specifically tailored to support the ARM program over the ARM ground-based observation sites. (Lines 142-143)"

"In this study, we used the cloud mask, cloud effective radius (r_e), cloud optical depth (τ), cloud liquid water path (LWP), cloud phase, and cloud top height variables in the product. We focus on warm boundary layer clouds with cloud top below 3km and a liquid cloud phase. The r_e and τ retrievals are based on the shortwave-infrared split window technique during the daytime. Cloud LWP is derived from r_e and τ using the equation: $LWP = \frac{4r_e\tau}{3Q_{ext}}$, where "$Q_{ext}$ represents the extinction efficiency and assumed constant of 2.0. Cloud mask algorithm is consistent with the CERES Ed-4 algorithm, as described in Trepte et al. (2019), where cloudy and clear pixels are distinguished based on the calculated TOA clear-sky radiance. Cloud top height is derived from the retrieved cloud effective and top temperature, together with the boundary-layer temperature profiles and lapse rate, as described in Sun-Mack et al. (2014). Cloud N_d is retrieved based on the adiabatic assumptions for warm boundary layer clouds, as described in Grosvenor et al. (2018) based on the following equation:

$$N_d = \frac{\sqrt{5}}{2\pi k}\left(\frac{f_{ad}c_w\tau}{Q_{ext}\rho_w r_e^5}\right)^{1/2} \quad (1)$$

In Equation (1), k represents the ratio between the volume mean radius and r_e, f_{ad} is the adiabatic fraction, c_w is the condensation rate, Q_{ext} is the extinction coefficient, and ρ_w is the density of liquid water (Grosvenor et al., 2018). (Lines 148-165)"

"We employed four one-way nested domains in the model, with the domain size of $27° \times 27°$, $9° \times 9°$, $3° \times 3°$, and $1° \times 1°$, and spatial resolution of 5km, 1.67 km, 0.56 km, and 190m, respectively, for d01, d02, d03, and d04 domain. The innermost domain (d04) exhibit a domain size close to most GCM grid spacing and is consistent with the spatial scale for quantification of cloud susceptibility in satellite study (e.g., Zhang et la., 2022, 2023; Qiu et al., 2024). The spatial resolution of 190 is much higher than the CPMs and close to the LES scale. All the analyses and evaluations in this study are based on output from the innermost domain (d04). (Lines 216-224) "

"To access the cloud responses to aerosol perturbations, we conduct three sets of simulations with different prescribed aerosol number concentration of N=100, 500, and 1000 cm^{-3} for all 11 cases. Cloud susceptibility is quantified as the change in domain-mean cloud properties within the innermost domain at the same output time, comparing polluted and clean conditions (e.g. N=1000 vs. N=100, N=500 vs. N=100, and N=1000 vs. N=500). With constant and uniform aerosol concentration, the N_d-LWP relations resulting from internal cloud processes are able to be quantified within each experiment at the same output time. To minimize N_d-LWP relations from cloud heterogeneity and small-scale covariability and to be consistent with the quantification of cloud susceptibility in satellite observations, the pixel level model outputs are smoothed to 25-km resolution and N_d-LWP relations are quantified as $dln(LWP)/dln(N_d)$ using the smoothed data. (Lines 269-279)"

- 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.

Thank you for the clarification and suggestion. We have changed the naming of different simulations as N=100, N=500, and N=1000 throughout the manuscript.

- 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.

Thank you for the question and constructive suggestion. We compared the LWP susceptibility and the occurrence frequency of different cloud states between the "high-ridge" and "post-trough" regimes (Figures R2), only one case is available in the "weak-trough" regime (Table S1).

In our previous study using six-year ground-based observations at the ARM ENA site, Zheng et al. (2025) found that the "high-ridge" regime has significantly more single-layer stratocumulus clouds, thinner cloud depth, smaller LWP, and smaller surface rain rate compared to the "post-trough" regime. Consistently, more non-precipitating thin clouds occur in the high-ridge regime compared to the post-trough regime, with the total frequency of occurrence of 49% and 40%, respectively in the simulations (Figures R2b and d). For cloud susceptibility, the non-precipitating thin clouds in the high-ridge regime exhibit more negative LWP susceptibility compared to clouds with similar LWP and Nd in the post-trough regime, likely due to the cold dry air above clouds with the subsidence in the high-ridge regime. Additionally, the non-precipitating or slightly drizzling thick clouds in both regimes exhibit strong positive LWP susceptibilities, indicating that the model-observation discrepancy for this cloud state is consistent with different synoptic conditions and warrant further investigations in the next section.

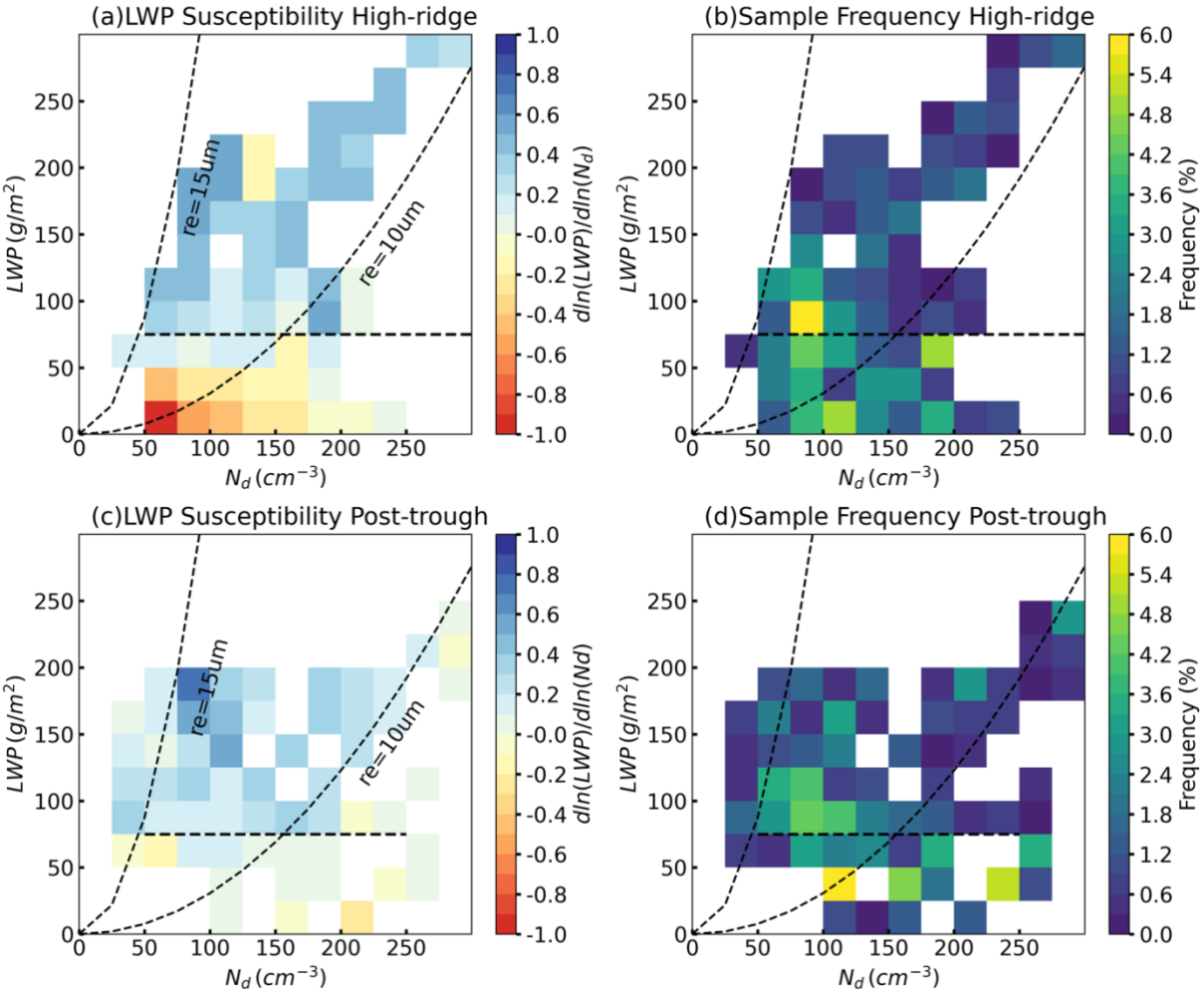

Figure R2. Mean liquid water path (LWP) susceptibility from WRF simulations for (a) (b) the high-ridge regime and (c) (d) the post-trough regime. (a) (c) cloud LWP susceptibility $dln(LWP)/dln(N_d)$, (b) (d) frequency of occurrence of sample in each bin.

To address this comment, we have added Figure R2 to the supplementary material and the following text to the manuscript:

"To further examine whether these discrepancies depend on large-scale meteorological conditions, we assessed LWP susceptibility across different synoptic regimes. Because only one case is available for the "weak-trough" regime (Table S1), our comparison focuses on the "high-ridge" and the "post-trough" regimes (Figure S10). The "high-ridge" regime shows a higher occurrence of non-precipitating thin clouds than the "post-trough" regime, with total frequencies of 49% and 40%, respectively (Figures S10b, d). This more frequent non-precipitating thin cloud in the model is consistent with our previous study based on six years of ground-based observations at the ARM ENA site, which revealed that the "high-ridge" regime favors single-layer stratocumulus clouds with shallower cloud depth and smaller LWP compared to the "post-trough" regime (Zheng et al., 2025).

In addition, non-precipitating thin clouds in the "high-ridge" regime exhibit more negative LWP susceptibilities than clouds with similar LWP and N_d in the "post-trough" regime. This difference in LWP susceptibility is associated with the colder and drier air above clouds under subsidence in the "high-ridge" regime, which facilitates cloud dissipation, as also demonstrated in the case study. Furthermore, non-precipitating or lightly drizzling thick clouds in both synoptic regimes manifest strong positive LWP susceptibilities, suggesting that the model-observation discrepancy for this cloud state persist regardless of synoptic conditions and therefore warrants further investigation. (Lines 566-583)"

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.

Thanks for the suggestion. The captions of the figures have been edited accordingly.

- 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.
Thanks for this insightful suggestion. We agree that Section 3.3.1 includes a lot of information of clouds from both observations and model outputs while clouds were further categorized by re and optical depth. This complexity sometimes makes the discussion difficult to follow. After attempting to reorganize the figures as you suggested, we found that this organization would split the discussion of model parameterization issues into three separate parts. For example, in the figure of clouds with $\tau < 10$, the issue of overestimation of precipitation at cloud top for thin clouds is reviewed and discussed. But the issue of overestimation of rain in thick clouds were not discussed until the figure of clouds with $\tau > 20$. Compared to the suggested arrangement, the current arrangement has the advantages of 1) combining the discussion of model issues and improvements together. 2) all the cloud characteristics shown in observations were combined together and discussed first, then they were compared to model simulations. As a result, we have decided to keep the current organization.

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.*

Thanks for the suggestion. We agree with the reviewer that the word "reconcile" is not accurate enough. The title has been updated to "Understanding the causes of Satellite–Model Discrepancies in Aerosol–Cloud Interactions Using Near-LES Simulations of Marine Boundary Layer Clouds"

- *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.*

Thanks for the suggestion. This sentence has been updated to "Aerosol–cloud interactions (ACI) remain the largest source of uncertainty in model estimates of anthropogenic radiative forcing, primarily because of deficiencies in representing aerosol–cloud microphysical processes that lead to inconsistent cloud liquid water path (LWP) responses to aerosol perturbations between observations and models."

- *Line 25: "a modest LWP decrease" to an increase in Nd.*

Edited.

- *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.*

These sentences have been edited to "Non-precipitating thin clouds exhibit a modest LWP decrease with increasing N_d (mean susceptibility = -0.13), consistent in sign but weaker in magnitude than satellite estimates due to enhanced turbulent mixing and evaporation. Meanwhile, the largest model-observation discrepancy occurs in non-precipitating thick clouds, where simulated LWP susceptibilities are strongly positive while observations indicate large negative values ($+0.32$ vs. -0.69)."

- *Line 108: please define the abbreviation of "MBL".*

Done.

- *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 Nd 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.*

Thanks. Cloud retrieval method for each variable used in the SEVIRI Meteosat cloud retrieval product has been added. We moved the equations for calculating N_d to here. Please refer to our reply to major comment #2.

- *Line 138: How was the satellite retrieval smoothed to 25-km resolution?*

The pixel-level satellite retrievals with a spatial resolution of 3km for Meteosat11 and 4km for Meteosat10 are averaged in each $25km \times 25km$ box to get the 25-km smoothed value. The 25-km cloud fraction is defined as the fraction of cloud pixels to the sum of cloudy and clear pixels in each box. As suggested by Feingold et al. (2022), N_d is retrieved at pixel level and then smoothed to 25 km.

- *Line 155: "0000 UTC"*

Modified.

- *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.*

Thanks for the suggestion, the subtitle of this section has been edited to "Datasets"

- *Lines 180-182: What are the spatial resolution of the other two nested domains?*

The spatial resolution of d02 and d03 domains are 1.67 km and 0.56 km, respectively. The following text has been added to the manuscript:

"We employed four one-way nested domains in the model, with the domain size of 27° × 27°, 9° × 9°, 3° × 3°, and 1° × 1°, and spatial resolution of 5km, 1.67 km, 0.56 km, and 190m, respectively, for d01, d02, d03, and d04 domain. The innermost domain (d04) exhibit a domain size close to most GCM grid spacing and is consistent with the spatial scale for quantification of cloud susceptibility in satellite study (e.g., Zhang et la., 2022, 2023; Qiu et al., 2024)."

- *Line 186: How often is the lateral boundary condition updated?*

The lateral boundary conditions are updated every three hours.

- *Lines 189-191: How are boundary layer and clouds treated in the innermost domain?*

In the innermost domain, with the spatial resolution of 190m (close to LES resolution), the boundary layer processes and shallow cumulus clouds are resolved. We turned on the PBL scheme and shallow cumulus scheme in the d01 and d02 domains, where the Mellor–Yamada–Janjic (MYJ; Mellor and Yamada, 1982) PBL scheme and the shallow cumulus schemes (Hong and Jiang, 2018) are utilized. In d03 and d04 domains, these processes are resolved.

- *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?*

The Meteosat cloud coverage is defined as the fraction of cloudy pixels to the summation of cloudy and clear pixels. The cloud mask algorithm used in Meteosat cloud retrieval product is consistent with the CERES Ed4 cloud mask algorithm described in Trepte et al. (2019), where cloudy and clear pixels are distinguished based on the calculated TOA clear-sky radiance for different surface conditions, time, viewing and illumination conditions. The cloud coverage in WRF simulation is similarly a spatial fraction of cloud in the domain, where cloudy and clear of each pixel is estimated based on relative humidity and cloud water mixing ratio. As a result, the comparison of domain cloud coverage between Meteosat and WRF model is a consistent evaluation of model performance.

The cloud top height retrieval in Meteosat product is based on the cloud top temperature along with the temperature profiles and lapse rate, with an error range of 0.04 and 0.1 km over ice-free water during daytime and night time, respectively (e.g., Sun-Mack et al., 2014; Minnis et al., 2021). Cloud top height in model output is based on the "CTOPHT" variable in the model, which is estimated as the highest model level where the cloud water mixing ratio exceed a threshold. The time series of Nd has been added to Figures 1 and 3 and the corresponding discussion has been added to the text. To address this comment, the Meteosat cloud mask and cloud top height retrieval algorithms were added to the data and methodology section. Please refer to the previous major comment #2 for the added text.

- *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.*

Thanks. This sentence has been edited as "In the N=100 simulation, WRF model reproduces the overcast and precipitating stratocumulus clouds, with a domain mean cloud cover varies

between 0.90 to 0.94 from 00-13 UTC, which is slightly below that from Meteosat of 0.97 to 1.0 (Figure 1a, blue and black lines)".

- *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.*

Done.

- *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?*

Thanks for your question. The lack of cloud dissipation and diurnal variation in marine boundary layer (MBL) clouds in the WRF model is likely associated with the biases in the thermodynamic profiles inherited from ERA5. As seen in Figures R1 left figure, on 21 July 2016, ARM sounding observations indicate a sharp decrease in moisture above the PBL between 14 and 20 UTC, leading to the dissipation of clouds after 14 UTC (Figure 1a). In contrast, both ERA5 and WRF simulation show a gradual decrease in specific humidity and relative humidity above the PBL from 0 to 20 UTC, resulting in a much moister layer above clouds in the model (Figures R3 middle and right). Consequently, clouds did not dissipate in the afternoon in the simulation.

On 25 July 2016, the ARM sounding observations similarly exhibit a pronounced decrease in specific humidity and relative humidity above the PBL between 14 and 24 UTC (Figure R4). In this case, the WRF simulation accurately capture the observed feature, reproducing a sharp decrease in moisture above the PBL from 14 to 24 UTC. As a result, clouds in the N=100 and N=1000 simulations dissipate from 14 to 24 UTC, consistent with satellite observation (Figure R5). This pattern also holds for other cases in the high-ridge regime, such as 22 and 28 July 2016, where the accuracy of the simulated PBL moisture variation determined whether the model captured the observed diurnal evolution of clouds (figures not shown). These cases demonstrate that the diurnal cycle of cloudiness is highly sensitive to the representation of diurnal variation in moisture as well as the moisture gradients near the inversion.

The fixed, vertically uniform aerosol concentration further contributes to the persistence of clouds by maintaining unrealistically high CCN concentrations throughout the day and suppressing precipitation. The lack of precipitation scavenging also reduces evaporative cooling and weakens cloud–PBL decoupling, inhibiting afternoon cloud breakup.

We added these discussions to the second last paragraph in Section 3.1:

"The absence of afternoon cloud dissipation in WRF simulations are likely associated with model biases in the thermodynamic structure inherited from ERA5. For example, on 21 July 2016, ARM sounding observations show a pronounced decrease in specific humidity and relative humidity above the PBL between 14 and 20 UTC (figures not shown). This sharp drying leads to cloud erosion in the observations. However, WRF simulations or ERA5 reanalysis produces only a gradual reduction in moisture from 00 to 20 UTC (Figure 2a), maintaining a moist layer above cloud top and prevent cloud breakup. On 22 July 2016, the model reproduces the moisture gradient above PBL with a warm and dry layer above, the lifted cloud top in the N=1000 simulation entrain dry air into cloud system and dissipate clouds in the afternoon (Figure 3a). On days when ERA5 accurately capture the observed moisture decrease above PBL (e.g., 25 and 28 July 2016), the model reproduces both the dissipation and evening redevelopment of clouds seen in Meteosat data (figures not shown). This indicates that the diurnal evolution of MBL clouds is highly sensitive to the representation of diurnal variation in moisture as well as the moisture gradients near the inversion.

The prescribed, vertically uniform aerosol concentration further reinforces cloud persistence by maintaining elevated CCN levels and suppressing drizzle formation. The lack of precipitation scavenging prevents cloud-base evaporative cooling and inhibits decoupling, both of which would otherwise promote afternoon cloud breakup. The implications of thermodynamic and aerosol-related biases for the estimated ACI are discussed in detail in Section 3.3.2. (Lines 436-455)"

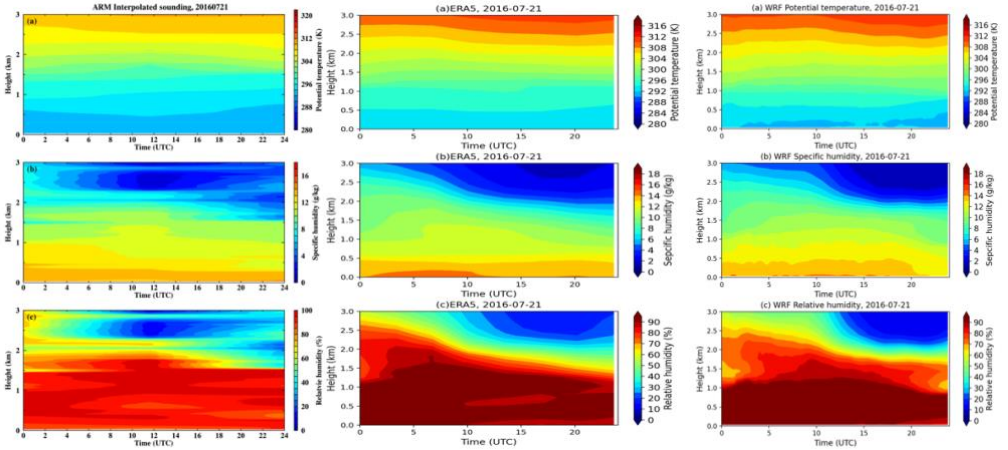

Figure R3 Time series of thermodynamic profiles on 21 July 2016, for (a) potential temperature (unit: K) (b) specific humidity (unit g/kg), (c) relative humidity in (left) ARM interpolated sounding, (middle) ERA5 reanalysis, and (right) in WRF N=100 simulation.

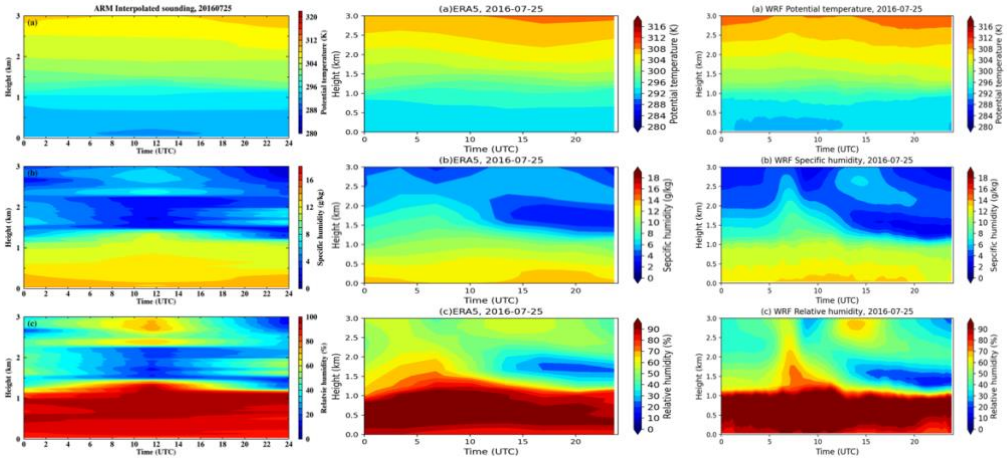

Figure R4 Time series of thermodynamic profiles on 25 July 2016, for (a) potential temperature (unit: K) (b) specific humidity (unit g/kg), (c) relative humidity in (left) ARM interpolated sounding, (middle) ERA5 reanalysis, and (right) in WRF N=100 simulation.

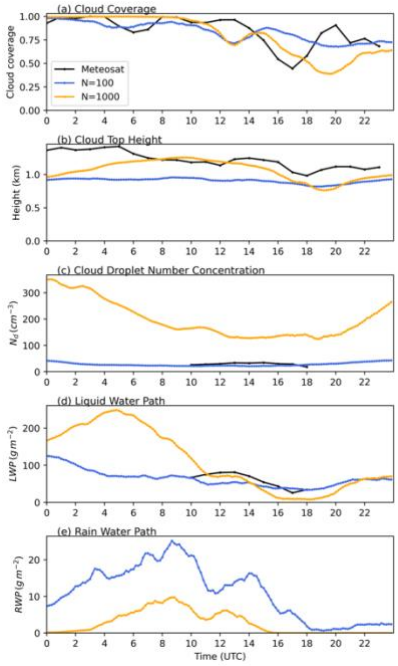

Figure R5. Time series of domain-averaged cloud properties from observations and model simulation on 25 July 2016. (a) Cloud coverage, (b) cloud top height, (c) cloud liquid water path, and (d) rain-water path for N=100 (blue lines) and N=1000 (orange lines) experiments.

- *Figure 4: Please use a full caption here.*

Done

- *Figure 5: "WRF simulations" are these from all polluted versus clean simulations or just one of the sets? Are Re 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.*

Figure 5 has been updated with same axis ranges for all subplots. Cloud susceptibility from WRF simulations are estimated based on all three aerosol concentrations between clean and polluted experiments (e.g. N=1000 vs. N=100, N=500 vs. N=100, and N=1000 vs. N=500) to estimate the mean LWP response. In Figure 5, the r_e =15 μm, r_e =10 μm isolines are estimated based on the relationships between r_e and LWP and r_e and N_d in the satellite retrievals: $LWP = \frac{4r_e\tau}{3Q_{ext}}$ and $N_d = \frac{\sqrt{5}}{2\pi k}(\frac{f_{ad}c_w\tau}{Q_{ext}\rho_w r_e^5})^{1/2}$. In Figure 6c, we show the WRF simulated mean re in the LWP-N_d space.

To address this comment, we added the following explanations on the LWP-N_d space and r_e isolines in the manuscript:

"Based on the relationships between r_e, LWP, and N_d in the satellite retrievals (e.g., $LWP = \frac{4r_e\tau}{3Q_{ext}}$, $N_d = \frac{\sqrt{5}}{2\pi k}(\frac{f_{ad}c_w\tau}{Q_{ext}\rho_w r_e^5})^{1/2}$), r_e =15 isolines is marked in the LWP-N_d parameter space as an commonly used indicator of precipitation likelihood in the satellite retrieval (e.g., Gryspeerdt et al., 2019; Toll et al., 2019; Zhang et al., 2022; Qiu et al., 2024). Based on the distinct LWP, cloud albedo and CF susceptibilities, MBL clouds are classified into three states: the precipitating clouds (r_e >15 μm), the non-precipitating thick clouds (r_e <15 μm, LWP> 75 gm^{-2}), and the non-precipitating thin clouds (r_e <15 μm, LWP< 75 gm^{-2}) (Qiu et al., 2024). To be consistent with satellite observations, clouds in WRF simulations are classified using the same definition. (Lines 479-488)"

- *Line 382: How does the Meteosat LWP susceptibility calculated?*

We added the following text to the data and methodology section to clarify the calculation of LWP and CF susceptibilities in Meteosat data: "In the context of ACI: cloud susceptibility quantifies how sensitive a cloud property responds to change in aerosol concentration or N_d. To constrain the spatial-temporal variation in meteorological conditions and cloud properties, cloud susceptibility is estimated as the regression slope between N_d and cloud properties within the 1° × 1° domain at each time step of satellite observations. Because of the non-linear relations between LWP and N_d, the LWP susceptibility is quantified in logarithm scale as $dln(LWP)/dln(N_d)$ (e.g., Gryspeerdt et al. 2019; Qiu et al., 2024), whereas cloud fraction (CF) susceptibility is quantified as $dCF/dln(N_d)$ (e.g., Kaufman et al. 2005; Chen et al., 2022; Qiu et al., 2024)."

We also added the calculation of LWP susceptibility in Meteosat observation in Figure 5: "To evaluate model simulation, LWP susceptibility from satellite retrievals is estimated within the same domain as the model configuration for the same 11 cases (Figures 5 c, d). More specifically, LWP susceptibility is estimated as the regression slope between LWP and N_d within the 1° × 1° domain at each time step of satellite observations. For precipitating clouds, LWP slightly decreases with aerosol perturbations in satellite data (Figure 5c). "

- *Line 386: What does it mean by "to be consistent with satellite observations"?*

This paragraph has been edited, please see the previous comment.

- *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.*

Thanks for the suggestions. The definitions of different cloud states have been added to the manuscript, please see the previous comment.

- 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?

Thanks for the question. This sentence in the paper was not accurate enough. Our satellite LWP susceptibility based on the selected 11 cases agrees well with the LWP susceptibility base on four years of satellite observations in our previous study (Qiu et al., 2024; Figure 2). As seen in Figure R6a, the LWP susceptibility is positive for precipitating thick clouds in ultra clean conditions with $N_d < 30\ cm^{-3}$ and LWP > 125 gm^{-2}. For most of the precipitating clouds, their LWP susceptibility is negative, which is consistent with the LWP susceptibility in this study for clouds with similar properties. The slight decrease in LWP for precipitating clouds is likely due to the depletion of LWP from the sedimentation–evaporation–entrainment feedback.

We have edited this sentence in the paper as follow: "For precipitating clouds, LWP slightly decreases with aerosol perturbations in satellite data, which is consistent with the LWP susceptibility derived from four years of data in the ENA region in our previous study (Qiu et al., 2024). This decrease of LWP with increasing N_d is likely associated with the depletion of LWP through sedimentation–evaporation–entrainment feedbacks, which outweigh the increase of LWP from precipitation suppression. In contrast, in model simulations, the lack of realistic evaporation-entrainment feedback results in LWP increasing primarily through precipitation suppression. The simulated LWP susceptibilities are significantly different with satellite observations at 95% confidence level for most precipitating clouds (Figure 5a). (Lines 520-528)"

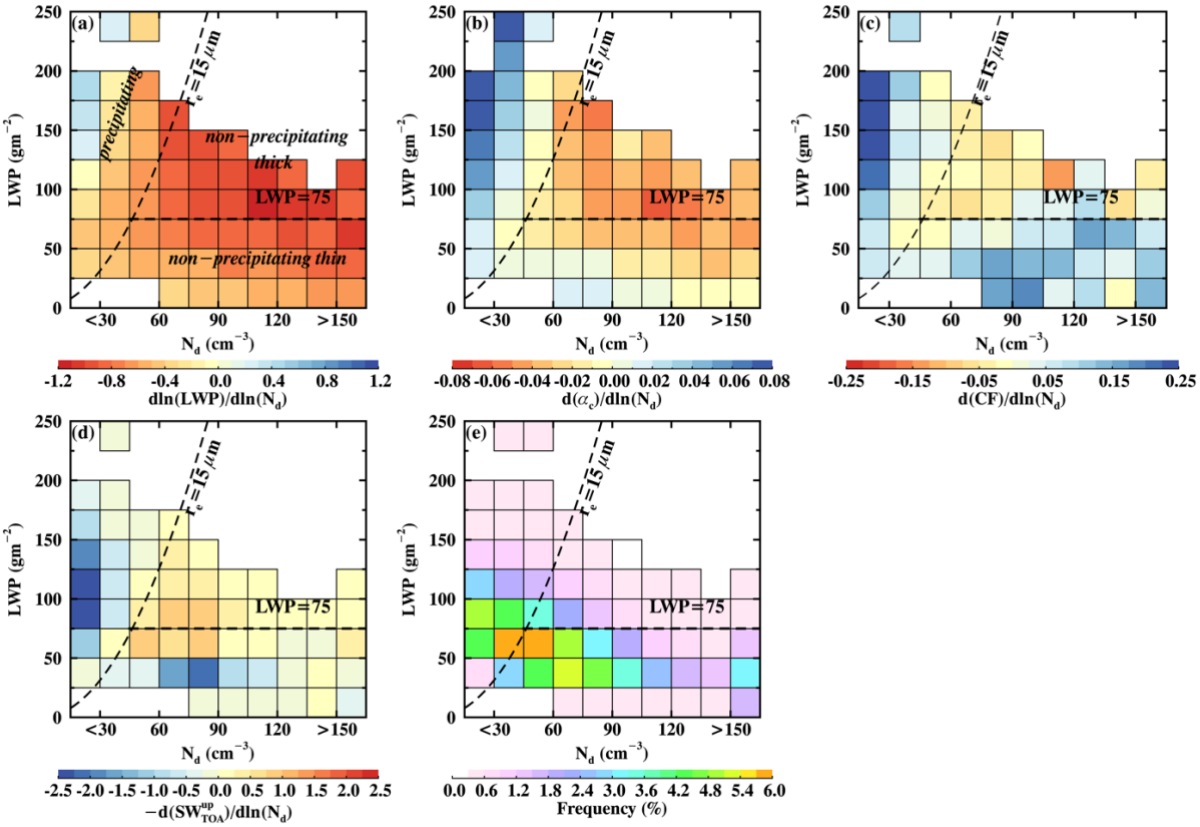

Figure R6. Mean cloud susceptibilities for different N_d and LWP bins during the daytime. (a) cloud LWP susceptibility ($dln(LWP)/dln(N_d)$), (b) cloud albedo susceptibility ($d\alpha_c/dln(N_d)$), (c) cloud fraction susceptibility ($dCF/dln(N_d)$), (d) cloud shortwave susceptibility ($-dSW^{up}_{TOA}/dln(N_d)$) weighted by the frequency of occurrence of samples of each bin, and (e) frequency of occurrence of samples in each bin. The dashed lines in (a)-(e)

indicate r_e =15 μm and LWP= 75 gm^{-2}, as thresholds for precipitation (precipitating clouds located to the left of the line) and thick clouds (with LWP > 75 gm^{-2}). The defined three clouds states are noted in (a). (Figure was adapted from Qiu et al., 2024)

- *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).*

Yes, I agree with you that this sentence was not accurate, and it has been edited.

- *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.*

Thanks for the clarification. This sentence has been modified. The agreements with previous model results indicate consistency instead of robustness of our model results.

- Figure 6: *It is confusing here that the r_e dashed lines across different r_e contour colors in (a) and (c). Please be clear about how each effective radius is calculated or derived in (a), (c) and the dash line.*

Thank you for pointing out this unclear point. In satellite observations, both LWP and N_d are retrieved as a function of function of r_e and τ . In Figure 6a, the r_e dashed lines are based on the relationships between r_e, LWP, and N_d in the satellite retrievals (e.g., $LWP = \frac{4r_e\tau}{3Q_{ext}}$, $N_d = \frac{\sqrt{5}}{2\pi k}(\frac{f_{ad}c_w\tau}{Q_{ext}\rho_w r_e^5})^{1/2}$), and are used as indications of precipitation in the figure. As a result, the r_e dash lines agree with the r_e contour in Figure 6a. To use the same classification of precipitation in the model as the satellite observations, in Figure 6c, the r_e dashed lines are based on the same relationships between r_e, LWP, and N_d in the satellite retrievals, while the r_e contour is from model output. As model underestimate r_e compared with satellite observations, the dash lines cross different contour colors as in Figure 6a.

- *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?*

Thanks for pointing out the mistake in the frequency of satellite data. We have corrected the mistake in the calculation of frequency, the occurrence frequency for different cloud states have been updated to 22.2%, 55.6%, and 22.2%, respectively. The occurrence frequency of precipitating, non-precipitating thin and non-precipitating thick clouds based on the selected cases align well with the occurrence frequency of cloud states in ARM data based on six-year of observations, with non-precipitating thin clouds the dominate cloud type and thick clouds the least frequent cloud type, suggesting that the selected cases are representative of the typical distribution of MBL cloud types in the ENA region in summer.

- *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.*

Thanks for the suggestion and clarification. The clean condition here indicates under clean environment in the observations when cloud r_e is greater than 15 μm. The naming of model simulations has been edited to N=100, N=500, and N=1000 in the manuscript instead of using clean and polluted.

- *Lines 506-507: "likely due to mixing and evaporation" – can you be more specific on this?*

As seen in Figure 7h, the maximum frequency of radar reflectivity shows a decrease of signal towards cloud base, which is likely due to the mixing and entrainment from cloud base which lead to the evaporation of cloud and rain drops.

- *Figures 8-10: Please use full captions for these figures.*
- Done
-

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.

Thanks for the suggestion. As we don't have the observed DSD data from filed campaign for most of our cases, I compare the DSD in one of our case with the measured DSD in Yeom et al. (2022) study based on the ACE-ENA campaign data on a different day.

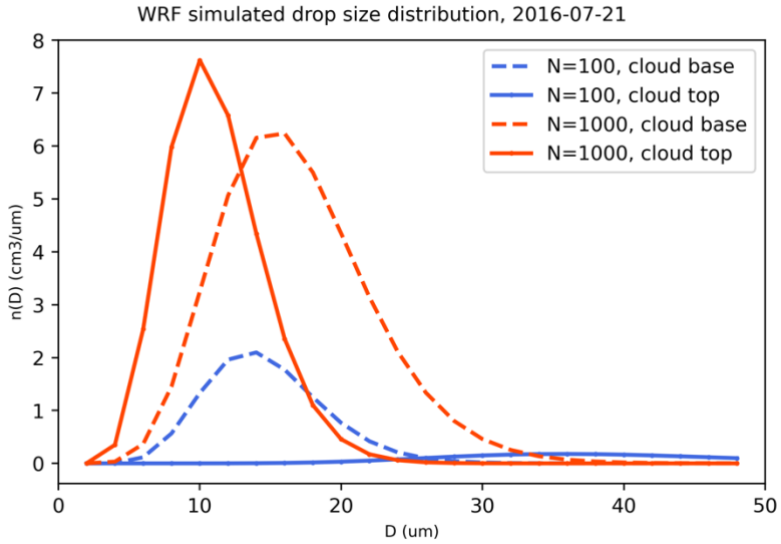

Figure R7: Cloud drop size distribution from WRF simulation on 21 July 2016.

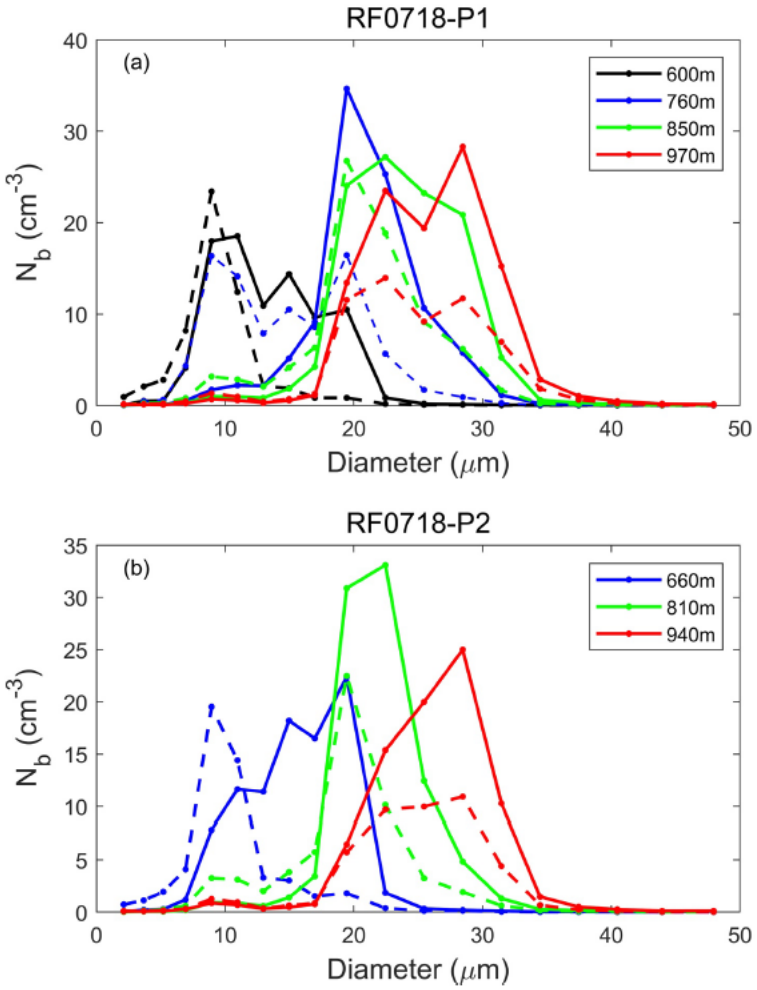

Figure R8 Average drop size distribution of bin number concentration of undiluted (solid lines) and diluted (dash lines) section at each horizontal penetration for (a) P1 (b) P2 flight on 18 July 2017 during the ACE-ENA field campaign. (Figure adopted from Yeom et al., 2022)

As shown in Figure R7, the DSD from N=1000 simulation is wider than that from N=100 simulation. This result is consistent with the parametrization of DSD in the Morrison scheme, where the dispersion parameter is proportional to N_d. As seen from Figures R7 and R8, the simulated DSD is wider than observation, especially for the N=1000 simulation.

The DSD of the model parameterization is one of our hypotheses to explain the excessive rain production in clouds with $r_e > 15\ \mu m$. Our results between different simulations show consistent characteristics with the DSD parameterization in the Morrison scheme, which support this hypothesis. For example, the overestimation of rain is more and more sever in the N=500 and N=1000 simulations comparing with the N=100 simulation for clouds with $r_e > 15\ \mu m$. However, as most of our cases don't have direct field campaign data, it is difficult to validate the DSD in the model.

- 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.

Thanks, this part has been moved to the method section.

- 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?

Following your suggestion, we calculated the domain mean cloud top height in WRF simulation based on the radar reflectivity profiles from the radar simulator and use the same reflectivity threshold to retrieve cloud top height as in ARM observation (reflectivity > -40 dBZ). Figure R9 compares the PDF between the cloud top height retrieved using reflectivity > -40 dBZ with that retrieved using cloud water mixing ratio > 0.001 g/kg for all 11 cases and all three aerosol concentrations. As seen in Figure R9, the retrieved cloud-top height is consistent between the two methods, with a difference in mean value of less than 40m. Therefore, the large discrepancy in cloud top RH between WRF simulations and ARM observations are not due to the different cloud top retrieval methods.

We have added the related discussion to the paper: "We further compare the cloud-top heights in WRF simulations defined using cloud water mixing ratio and radar reflectivity profiles with Z_e > -40 dBZ. The two approaches yield nearly identical results, with a mean difference of less than 40m (figure not shown). (Lines 855-858)"

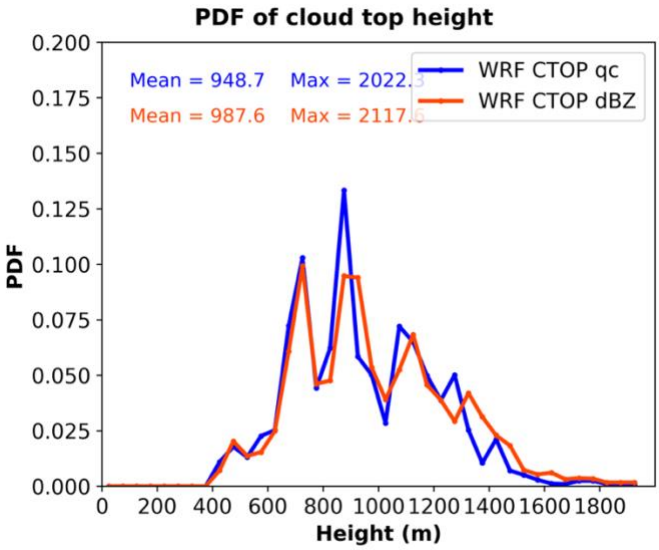

Figure R9. PDF of WRF simulated cloud top height retrieved using cloud water mixing ratio threshold (blue line) and using cloud radar reflectivity threshold (red line).

- *Lines 644-648: I recommend 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?*

Thanks for the question. The spatial and temporal representation issues is resolved by calculating the model output over a 10km × 10km grid box centered at the ARM ENA site for each sounding time. Given the ~1.2-1.4 km mean cloud-top height for MBL clouds at ENA during summer, and the balloon sounding rise at ~1m/s speed, it takes ~1200s-1400s to reach the cloud top. With the prevailing wind speed of 7 m/s, the balloon travels ~8 to 10km horizontally. Therefore, we averaged the WRF pixel-level output over a 10km × 10km grid box centered at the ARM ENA site for each sounding time (usually at 12:00 UTC and 0:00 UTC each day).

- *Figure 12: What is the shaded area for?*

The solid blue line is the median value for each RH bin and the shaded area shows the lower and upper 25th percentiles of cloud susceptibility in each cloud top RH bin. This information has been added to figure captions of Figures 12 and 13.

- *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)?*

Thanks for the clarification question. Yes, Figure 12 shows the mean relationships between cloud top RH and cloud susceptibilities calculated based on all simulations (e.g. N=1000 vs. N=100, N=500 vs. N=100, and N=1000 vs. N=500). The cloud-top RH is also based on all simulations using cloud-top heights in N=100, N=500, and N=1000 simulations. Figures R10 and 11 show these relationships between N=100 and N=500, N=100 vs. N=1000, respectively. As seen in these figures, the positive relationships between cloud-top RH and cloud susceptibility are consistent across different aerosol concentrations.

To address this comment, we have added the following text to the manuscript:

"Figure 12 shows the mean relationship between clout-top RH and cloud susceptibilities calculated based on domain mean values for all three simulations (e.g. N=1000 vs. N=100, N=500 vs. N=100, and N=1000 vs. N=500). The cloud top RH is the domain mean RH value at ~100m above cloud top for all simulations. As seen in Figure 12a, we find a positive correlation between cloud-top RH and LWP susceptibility in the simulations, which is consistent with cloud responses shown in case study where a dry layer above cloud promotes evaporation and decrease LWP. Additionally, these positive relationships are consistent among different aerosol concentrations (e.g., N=1000 vs. N=100 or N=500 vs. N=100; figures not shown). (Lines 870-878)"

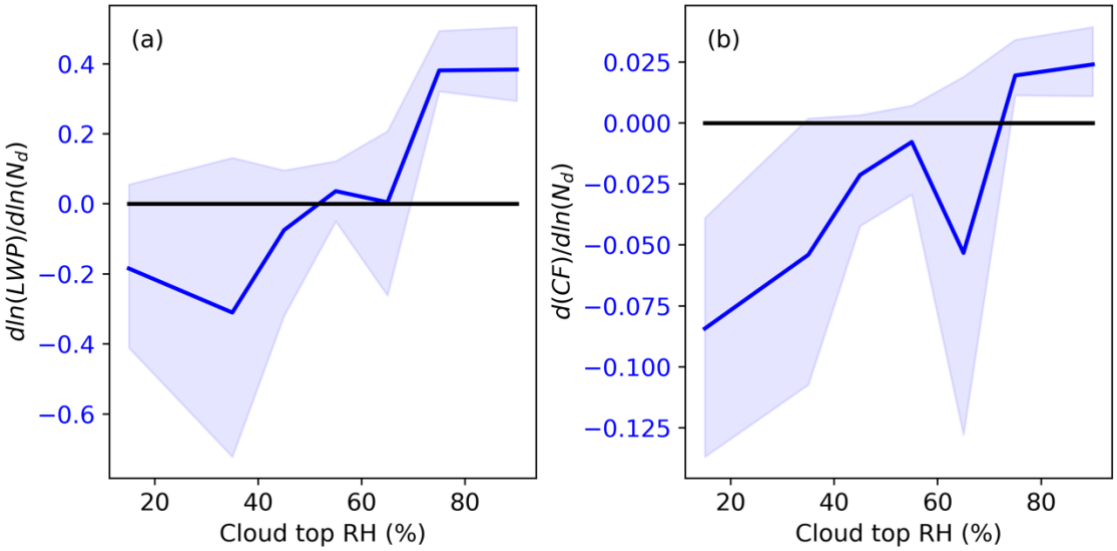

Figure R10. Dependence of (a) LWP susceptibility and (b) CF susceptibility on cloud top relative humidity in WRF simulations between N=100 vs. N=500. The solid blue line shows the median value of each RH bins and the shaded area shows the lower and upper 25th percentiles.

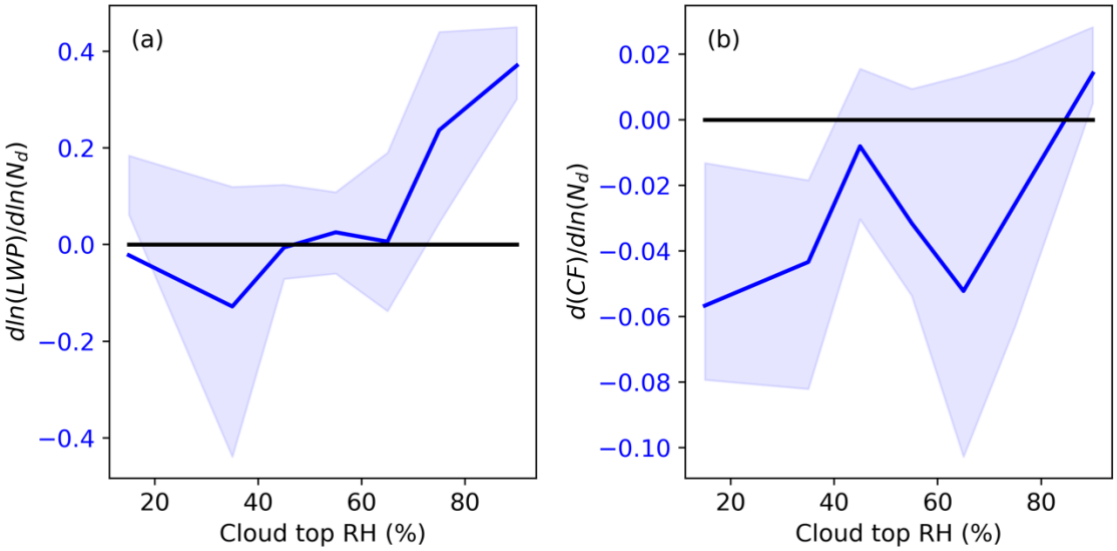

Figure R11. Dependence of (a) LWP susceptibility and (b) CF susceptibility on cloud top relative humidity in WRF simulations between N=100 vs. N=1000. The solid blue line shows the median value of each RH bins and the shaded area shows the lower and upper 25th percentiles.

- 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?
Figure 13 and the figure caption have been updated.

Figure 14: Are these from simulations with all different aerosol concentrations?
Yes, results shown in Figure 14 are based on the 11 cases and all three aerosol concentrations.
- Figure 15: Please use a full caption.
- Done.
- Lines 745-754: I think several references are missing here in this first paragraph when mentioning the findings from previous studies.
Thanks. The references have been added to the manuscript.
- Lines 762-763: I suggest adding the LWP bias here rather than using "generally match".
Thanks. This sentence has been updated to "The simulated MBL clouds generally match the satellite observation in domain mean cloud coverage and mesoscale

organization (Figures 1, 3, S2-S4), while the model may struggle to capture the diurnal evolution of clouds, especially the dissipation of clouds in the afternoon. Model overestimate cloud LWP, especially in the polluted runs and underestimated cloud top height compared to satellite retrievals."

- 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?*

The overestimations of N_d are due to the overestimated prescribed aerosol concentration in model setting combined with the lack of precipitating scavenging effect in WRF model. The overestimation of LWP is likely due to the positive LWP susceptibility for thick clouds, as shown in Figure R1a. This sentence has been updated to "Meanwhile, the non-precipitating thick clouds are the dominant cloud state in the model, with a total frequency of 49%, compared to a 15.7% frequency of occurrence in satellite observations. The overestimation of N_d arise from the overestimated aerosol concentration in the configuration, combined with the absence of precipitation scavenging in the model. The overestimation of LWP is due to the positive LWP susceptibility in thick clouds where LWP in N=100 simulation show good agreement with satellite retrievals (Figure S9). (Lines 1034-1038)"