Responses to Reviewer 3

Dear Reviewer, we appreciate your time and effort in acknowledging and thoroughly reviewing our manuscript. We are truly grateful for your constructive comments and insightful suggestions, which encourage and help us to improve the manuscript. We have revised the manuscript carefully based on your comments.

In the responses below, your comments are provided in black text and our responses are provided in blue text.

This work uses a regime-based approach to investigate aerosol-cloud interactions, specifically LWP and Nd, in warm marine clouds. The data set incorporates 2006 to 2014 measurements from CALIPSO for aerosol properties with MODIS for cloud properties in comparison to E3SMv2 simulations with machine learning to cluster four synoptic regimes in the ENA. Overall, they find that model relationships match observations qualitatively, but have more sensitive relationships likely due to model representation of cloud processes.

The paper reads very well with very few typos. The authors do a good job presenting necessary background and what's at stake in the introduction, starting with ACI and how it is measured then how it is modeled and how the two disagree. I understand how difficult it is to compress many of these complicated concepts such as precipitation suppression or enhanced entrainment-induced evaporation which introduce uncertainty, but this paper would benefit from more explanation of these processes.

We sincerely appreciate your thoughtful and constructive feedback. All comments have been carefully considered, and the manuscript has been revised accordingly.

For the data and method and section, I think this paper would benefit from discussion of the satellite products chosen. The literature suggests that most satellites with same measurement types are in good agreement, but I'm curious about the omission of MODIS Terra as it would provide more temporal resolution. Figures 2 and 4 are great for showing that models qualitatively match Nd-LWP and aerosol extinction coefficient-Nd relationships in observations, but are more exaggerated. The comparisons between observations and model parameters is well summarized at the end of each section.

We restricted the satellite data set to the curated CALIOP–MODIS Aqua collocation introduced in Painemal et al. (2020) and Li et al. (2025). That product was designed around daytime MODIS Aqua cloud retrievals processed with the CERES Ed. 4.0 algorithms and paired with CALIOP-S aerosol extinction, providing consistent sampling and reduced retrieval artifacts.

Adding MODIS Terra would increase temporal sampling, but its morning overpassing time (~10:30 LT) does not coincide with the early-afternoon (~13:30 LT) overpassing time of CALIOP on CALIPSO. Hence our collocation specifically targets the afternoon local overpass to keep aerosol-cloud relationships at a single time of day. Combining Terra with Aqua would therefore introduce diurnal aliasing into the ACI relationships derived. Prior work shows that the aerosol indirect effect and related cloud adjustments vary over the diurnal cycle (Diamond et al., 2020; Smalley et al., 2024), so fixing the local time is critical for a clean evaluation.

Finally, the Painemal et al. (2020) collocation/aggregation strategy was built and quality-controlled for Aqua, and it demonstrably reduces N_d bias, replicating that framework for Terra would require additional harmonization. We are interested in building such morning database for future work.

We have added the associated discussion in the revised Section 2.1:

'Moreover, the aggregated collocation method significantly reduces the MODIS Aqua N_d bias (Painemal et al., 2020), resulting in a relationship between aerosol and cloud properties less affected by artifacts. Note that to avoid diurnal variations in aerosol-cloud relationships, we fix the sampling to the Aqua local-afternoon overpass and do not merge with Terra morning orbits, while extending the collocation and quality-control framework to Terra is left for future work.'

While very thorough, I can't help but feel as though the section 3's focus on seasonal comparisons between satellite and model parameters and relationships is unnecessary to this body of work, especially with the end of that section mentioning its limitations and the need to separate by regime to disentangle meteorological variability which I think is the meat of this body of work.

We appreciate the suggestion, we have now significantly streamline Section 3 to make the discussion concise, and compelling for the necessitate of meteorological-regime-based analysis of the ACI in Section 4.

The comparisons between observations and the E3SMv2 models in each regime are thorough and any discrepancies have a one or more hypothesis supported by the literature. The paper concludes with many suggestions of next steps on how to improve model representation of observations, leaning on the difficulties of overcoming satellite measurement uncertainty. I think it would be useful to sort of rank future changes that are most feasible or important to implement. I'm excited to see future work in other regions, especially of the number of regimes found and how these relationships change, especially with more drastic differences in sources of aerosol.

We appreciate the suggestion, and Review 1 also raised similar concerns. We have added the following discussion in the conclusion section:

'In terms of the feasibility of potential model improvements, we think that a feasible approach would be the fine-tuning of the microphysical parameterization, ideally constrained by high-resolution observational data from field campaigns such as ARM. This may reduce the persistent uncertainties in simulating aerosol-cloud interactions, particularly under the dynamic meteorological transitions typical of the ENA region. Furthermore, emulation from high-resolution modeling (e.g., LES) of cloud and rain microphysics processes can be used to replace the bulk microphysics scheme, which can contribute to better performance with manageable cost as shown in previous studies such as Gettelman et al. (2021). Increasing spatial resolution is also feasible in a regionally refined mesh, and increasing vertical resolution might follow, but both would noticeably increase computational cost, so trade-offs should be considered with caution. Lastly, the development of new schemes that bridge the gap between shallow and deep cloud regimes remains particularly challenging, as current large-scale model schemes still treat them separately.'

Specific comments:

L48 Nitpicky, but satellite remote sensing's main advantage over in-situ cloud measurements is the spatial extent. The temporal resolution is often worse, especially of polar orbiting satellites.

We have changed this statement to 'Satellite remote sensing observations are essential in efforts to quantify the cloud adjustment to aerosol perturbations, by providing spatially extensive datasets.'

L49 It would be good to explicitly cite more of the numerous papers advancing ACI using satellite data.

Following your suggestion, more papers on the satellite advance on the ACI are cited:

'Numerous studies using satellite data have demonstrated a significant relationship and progressively advanced our understanding of cloud adjustments to aerosols (Bellouin et al., 2020; Diamond et al., 2020; Yuan et al., 2023; Feingold, et al., 2025; Goren et al., 2025).'

L67 It would help to introduce the inverted V-shaped relationship observed in satellite retrievals in the previous paragraph. It is currently explained in detail much later in L283.

Thanks for the suggestion, the inverted V-shaped relationship is now introduced in detail in the introduction, rather than Section 3.

L159 It would help to include a typical range of relative error values in retrieval and r-values of comparisons if available to better quantify what is meant by significant and decent.

We have now included the error estimate from previous study as below:

'…previous studies have shown that the N_d compares well with measurements from 11 aircraft campaigns, demonstrating a decent correlation when sampling the marine stratocumulus clouds, with r^2 values of 0.5~0.8 (Gryspeerdt et al., 2022). Therefore, to minimize known retrieval uncertainties, we focus on low-level liquid clouds where satellite N_d shows the strongest aircraft agreement and typical normalized root mean squared deviation of ~30-50 % (Gryspeerdt et al., 2022).’

L242, I think figure 1 would benefit from a 5th column of annual statistics.

The annual statistics are now added as 5th column in Figure 1:

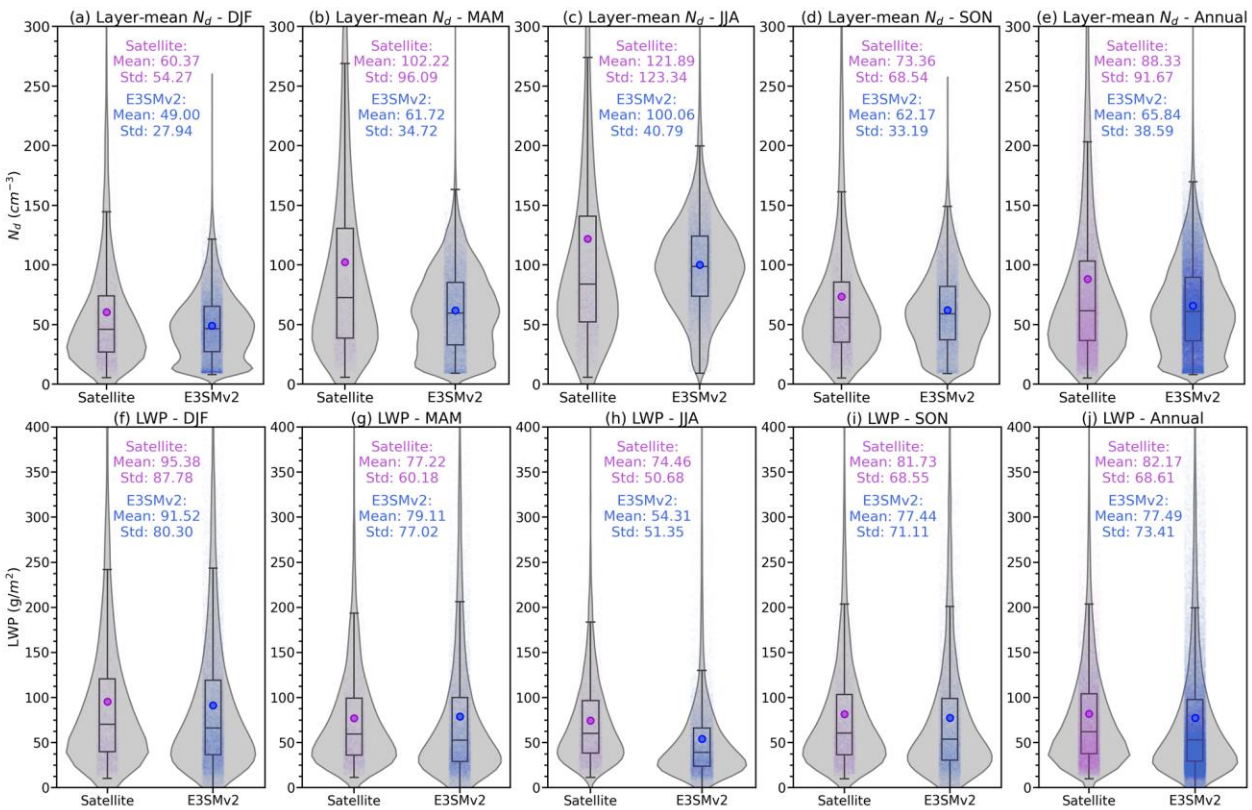

Figure 1. Violin plots of cloud droplet number concentration (N_d, top panels) and cloud liquid water path (LWP, bottom panels) from satellite retrievals (purple) and E3SMv2 simulations (blue), during winter, DJF (a, f), spring, MAM (b, g), summer, JJA (c, h), fall, SON (d, i), and Annual (e, j). The mean value is indicated by the color-coded dot. The smoothed shape of each violin shows the Gaussian kernel density estimate (KDE). From top to bottom within each violin, the box plot lines represent the third quartile (Q3, 75th percentile), median (Q2, 50th percentile), and first quartile (Q1, 25th percentile), respectively. The upper whisker extends to Q3 + 1.5 × IQR (interquartile range), and the lower whisker extends to Q1 − 1.5 × IQR.

And the corresponding description has been revised to:

‘Figure 1 illustrates the seasonal variations in the Nd and LWP for low-level clouds over the ENA, from satellite (MODIS) retrievals and E3SMv2. Annual means are 88.33 ± 91.67 cm^{-3} and 82.17 ± 68.61 g m^{-2} for satellite N_d and LWP, and 65.84 ± 38.59 cm^{-3} and 77.49 ± 73.41 g m^{-2} for E3SMv2 (Fig. 1e and 1j).’

L288 It would be helpful to provide ranges of Nd of these regimes found in previous work for comparison.

Thanks for the suggestion. We have included the previous reported N_d ranges as below:

‘In the satellite observations (Fig. 2a), increasing N_d suppresses precipitation, leading to LWP accumulation at low N_d. At higher N_d, the LWP response turns negative, consistent with enhanced entrainment and evaporative losses. These results are broadly consistent with prior satellite studies over marine stratocumulus: an inverted-V LWP-N_d relationship has been reported for N_d ranges of ~10-300 cm^{-3} in the southeast Pacific (Goren et al., 2025), globally (Gryspeerdt et al., 2019; 2022), and ~7-400 cm^{-3} for subtropical stratocumulus (Possner et al., 2020).’

L294 Missing space after Nd.

Thanks, it is corrected.

L332 The explanation of enhanced sea salt emissions with increased wind speed makes sense and should be observed globally, but is the seasonal dependency on the under and overestimation of the model aerosol extinction coefficient regionally dependent then if dust is a factor? Additionally, it would help to provide more context of where the transported dust in the ENA comes from.

The seasonal dependence of E3SM’s σ_{EXT} bias in the ENA is consistent with broader, global tendencies in the model’s dust cycle rather than a signal unique to ENA. In E3SM, dust vertical/long-range transport is generally underrepresented relative to CALIPSO (Feng et al., 2022), and regional dust AOD differences persist even when the global mean is constrained. Thus, the ENA seasonal pattern likely reflects these global dust process biases expressed through regional meteorology and sampling.

Overall, ENA dust originates from North Africa (Sahara Desert) and arrives episodically, the winter-spring synoptic events can drive northwestward intrusions that sporadically impact the ENA, and hence potentially impacts the seasonality on the aerosol extinctions (Logan et al., 2014; Rodríguez and López-Darias, 2024).

We have revised the associated discussion in Section 3.3 of the manuscript:

'The observed high σ_{EXT} in cold seasons reflects coarse-mode contributions from both enhanced sea-salt emissions under strong MBL winds and episodic Saharan dust intrusions that reach the ENA via the synoptic northwestward transport (Logan et al., 2014; Gläser et al., 2015; Rodríguez and López-Darias, 2024). E3SMv2 likely underpredicts this signal due to low sea spray (Burrows et al., 2020) and an underrepresentation of dust vertical extent and transport (Feng et al., 2022), a broader model tendency that appears over the North Atlantic as well (H. Wang et al., 2020; Qin et al, 2024).'

L365 It would be helpful to provide the sensitivity values from the cited studies for comparison.

We have added the following discussion:

'The positive ACI_N reflects the Twomey effect and lies within reported satellite ranges over marine stratocumulus and the eastern Atlantic. For example, McCoy et al. (2017) found a log-log slope of 0.31 between N_d and sulfate mass, Jia et al. 2021 reported 0.14-0.51 for N_d versus AOD over oceans, and recent reviews summarize satellite-based susceptibilities of about 0.1–0.7 depending on sampling and aerosol proxies (Gryspeerdt et al., 2023).'

L513 It would be helpful to provide a range of the peak of the inverted V-shape across the regimes in text to highlight this point.

We have added the following quantitative range of peaks:

‘Across regimes, the peak in the satellite data of the inverted LWP-N_d curve occurs at $N_d \approx$ 15-96 cm^{-3} (particularly, 96.5 cm^{-3} at Regime 4), and at LWP ≈ 101-136 g m^{-2}; whereas the E3SMv2 peaks at a much narrower $N_d \approx$ 15-19 cm^{-3} with higher peak LWP ≈ 142-171 g m^{-2}.’

L523 While I understand that $\mathscr{L}0$ is bulk, wouldn't separating by the peak of the inverted V-shape illustrate your points more of the rapid decrease?

We have computed the below table of the post-peak (N_d > peak binned N_d) log–log slopes for LWP vs N_d in each regime.

Regime	**Peak LWP corresponding binned N_d (satellite)**	**Peak LWP-N_d binned mean N_d (E3SM)**	**Slope ± SE from Sample > peak bin (satellite)**	**Slope ± SE from Sample > peak bin (E3SM)**
1 (Pre-Trough)	15.42	18.90	-0.179 ± 0.022	-0.377 ± 0.017
2 (Post-Trough)	18.90	15.42	-0.214 ± 0.020	-0.390 ± 0.017
3 (Ridge)	18.90	15.42	-0.176 ± 0.008	-0.327 ± 0.008
4 (Trough)	96.49 (last N_d bin)	15.42	N/A due to too few samples	-0.417 ± 0.042

In all four regimes, the post-peak slopes are negative for both datasets, and E3SM is consistently more negative (steeper decline) than satellite. This preserves the same sign and ordering as the bulk analysis, so the qualitative message is unchanged. Note that in regime 4 satellite has very few valid points and a large standard error since the LWP peak at the last N_d bin, which would invite distracting caveats if moved into the main text.

Therefore, as $\mathcal{L}_0$ is bulk metric by design. Mixing it with a subset-conditioned (post-peak) slope in the main text can blur the narrative and burden readers with extra conditioning details.

Figure 8 Nitpicky, but I think this figure would benefit from labels of the tickmarks between 10 and 50 as a significant part of the data is within that range.

The Figure 7 & 8 are replotted with more information on the x-axis tickmarks.