

Responses to Reviewer 2

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.

Review of “Regime-based Aerosol-Cloud Interactions from CALIPSO-MODIS and the Energy Exascale 2 Earth System Model version 2 (E3SMv2) over the Eastern North Atlantic” by Zheng et al.

In this work, the authors investigate aerosol-cloud interactions (ACI) in the Eastern North Atlantic (ENA) with satellite observations and nudged Energy Exascale Earth System Model version 2 (E3SMv2) simulations. In particular, the authors examine differences in liquid water path (LWP), droplet number concentration (Nd), and their covariance between observations and simulations. They find that, in general, there are systematic seasonal discrepancies between E3SMv2 and satellite observations of LWP and Nd that line up with prior studies. They also find the presence of the “inverted-V” in the models and observations, with a more pronounced V shape in the model output. To investigate the effect of ENA meteorology on these results, the authors employ a machine learning method to partition the data into 4 regimes based on synoptic conditions: pre-trough, post-trough, ridge, and trough. This more-targeted analysis reveals additional, interesting insights into the differences between the model and the observations and allow for more specific inferences on the importance of meteorology on ACI processes.

The paper is well-organized and detailed, showcasing valuable results that are interesting on their own merit and motivate exciting future research. The reanalysis-clustering method for regime analysis in particular I thought was a strong result with broad-reaching applications. ENA is strongly governed by transient synoptic weather systems, and I felt this methodology provided an interesting alternative to the usual methods of regime classification. While I think this is overall a

strong paper, I do have some comments that I feel should be addressed before publication. These are detailed below. Good work!

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

General comments

One key problem I have with this paper is that I think it “buries the lede” in terms of its subject matter. In the abstract, the regime names seem to come out of nowhere, and the title gives no indication that synoptic regimes are a key piece of subject matter for this work. Synoptic systems as an important lens through which we should be viewing this data aren’t suggested until well into the paper (end of section 2) and aren’t truly discussed until section 4. While I understand the desire to maintain a clear narrative, I think highlighting this portion of the analysis more clearly in the abstract and motivating it more directly in the introduction would help strongly with readability.

Thanks for the suggestion. We have carefully revised the abstract, introduction, and the narrative in Section 3 to better reflect the synoptic regime-based framework of the analyses.

The new abstract now reads:

‘Abstract. This study investigates aerosol-cloud interactions in marine boundary layer (MBL) clouds using an advanced deep-learning-driven synoptic-regime-based framework, combining satellite data (CALIPSO vertically resolved aerosol extinction and MODIS cloud properties) with 1° nudged Energy Exascale Earth System Model version 2 (E3SMv2) simulation over the Eastern North Atlantic (ENA; $\sim 10^\circ \times 10^\circ$, 2006-2014). The E3SMv2 captures observed seasonal variations in cloud droplet number concentrations (N_d) and liquid water path (LWP), though it systematically underestimates N_d . We then partition ENA meteorology into four synoptic regimes (Pre-Trough, Post-Trough, Ridge, Trough) via a deep-learning clustering of ERA5 reanalysis fields, enabling regime-dependent aerosol-cloud interactions analyses. Both satellite and E3SMv2 exhibit an inverted-V LWP– N_d relationship. In Post-Trough and Ridge regimes, the satellite shows stronger negative LWP– N_d sensitivities than in Pre-Trough regime. The Trough regime displays a muted

satellite LWP response. In comparison, the model predicts more exaggerated LWP responses across regimes, with LWP increasing too quickly at low N_d and decreasing more sharply at high N_d , especially in Pre-Trough and Trough regimes. These exaggerated model LWP sensitivities may stem from uncertainties in representing drizzle processes, entrainment, and turbulent mixing. As for N_d susceptibility to aerosols, N_d increases with MBL aerosol extinction in both datasets, but the simulated aerosol-cloud interactions appear oversensitive to meteorological conditions. Overall, E3SMv2 better captures aerosol effects under regimes that favor stratiform clouds (Post-Trough, Ridge), but performance deteriorates for regimes with deeper, dynamically complex clouds (Trough), highlighting the need for improved representations of those cloud processes in climate models.'

And we have highlighted the importance of synoptic regimes in assessing the ACI in the introduction:

'Moreover, synoptic systems organize boundary-layer clouds on multi-day timescales and strongly modulate aerosol-cloud-precipitation pathways (Mechem et al., 2018; Lee et al., 2025). Therefore, quantifying the untangled aerosol-cloud sensitivities require conditioning on the synoptic environment. For example, Zhang et al. (2022) found that the relationship between LWP and N_d is not only sensitive to aerosol loading but also modulated by the underlying meteorological conditions. And McCoy et al. (2020) used a cyclone compositing approach to demonstrate that aerosol-cloud interactions (e.g., the sign of LWP change with N_d) can differ inside vs. outside midlatitude cyclones. These considerations motivate our use of an objective synoptic-regime classification to control meteorology when evaluating the synoptic-regime-dependent ACI.'

In line with my previous comment, I feel the paper is a little lacking in terms of background on synoptic regime analysis with respect to ACI. While the authors do mention McCoy et al., 2020, which is important background for this work, it is done in the somewhat vague context of "atmospheric regimes" (particularly when "cloud regimes" and "meteorology regimes" are specified in the next few paragraphs). I think, generally, prior authors examining ACI in a synoptic

meteorology context have taken a more cyclone-specific approach (as in, compositing around low-pressure centers), as opposed to the trough/ridge classification approach. Differentiating synoptic regimes from more general “atmospheric” regimes here would help with seeding this idea early on. Also, I think adding some additional analysis contrasting these results with the prior literature in this area (e.g., McCoy et al., 2020) would add some critical context to the results/conclusions of this paper. If the authors feel that the analysis presented is too novel to be usefully compared to prior analyses of ACI in synoptic scale contexts, then that needs to be defended more thoroughly in the manuscript.

We appreciate the reviewer’s feedback and agree that referencing cyclone-based ACI literature strengthens our work. Our goal is to introduce a data-driven methodology to untangle meteorology from cloud responses, allowing the machine-learning framework to decide what aspects of the atmospheric pattern matter, without relying on pre-identifying specific synoptic systems (e.g., cyclones). This approach performs well: the clustered patterns naturally encompass the traditional cyclone separation, including pre-trough, post-trough, and trough phases. We also view this as transferable to other regions of the globe, including marine stratocumulus areas with weaker cyclone influence (e.g., the southeastern Atlantic).

That said, we have made several revisions to acknowledge and compare with prior work:

We have added references to McCoy et al. (2020) and related studies in the Introduction to highlight the importance of synoptic regime in aerosol-cloud interactions, as stated in above response. We acknowledge prior evidence that meteorological context influences ACI, thereby motivating our use of a synoptic clustering framework. This addition makes it clear to readers that our approach is built on the foundation of such findings, as we are extending the idea of regime-dependent ACI that others have established.

In Section 5 of the revised manuscript, we explicitly contrast our findings with McCoy et al. (2020). They reported LWP increases with aerosol in cyclonic regimes and little or opposite effect outside cyclones; we see the same pattern in satellite observations, enhanced LWP at higher N_d in the Pre-Trough (cyclone) regime and suppressed LWP in the Ridge (non-cyclone) regime. We also note

that our model evaluation aligns with McCoy's diagnosis of climate model bias: just as UM GA7.1 overestimated LWP response outside cyclones, E3SMv2 tends to over-respond in the Ridge regime. We qualitative argue that our study's conclusions support and extend the conclusions of cyclone-focused studies, thereby firmly embedding our contributions in the context of existing literature, while offering a more flexible, generalizable pathway via objective synoptic clustering.

The added discussion now reads:

'It is noteworthy that previous ACI studies in a synoptic context have been largely cyclone-centric (e.g., McCoy et al., 2020; Lee et al., 2025). Our regime-stratified results are consistent with that literature: in cyclone-associated conditions (Pre-Trough, Trough) we see LWP increases or smaller decreases with higher Nd, whereas in the anticyclonic conditions (Ridge) LWP decreases markedly with higher Nd, as expected in stable, dry high-pressure environments. Our clustering approach extends the synoptic pattern classification by providing a flexible, data-driven identification that captures the same physical contrasts as cyclone masks while explicitly considering the other two regimes (Ridge and Post-Trough). Therefore, our approach might be more general, and remaining applicable beyond the regions dominated by midlatitude storm tracks. Model behavior also parallels prior findings: E3SMv2's overly steep LWP reductions in Ridge conditions mirror the overestimation of LWP sensitivity outside cyclones reported by McCoy et al. In short, our aim is to develop a data-driven way to untangle meteorology from cloud responses without pre-specifying synoptic systems, and the learned regimes would naturally recover the traditional cyclone phases (pre-, post-trough, trough). And we also view this as transferable to other regions of the globe, including marine stratocumulus regions with weaker cyclone influence (e.g., the southeastern Atlantic).'

Throughout the manuscript, there are many figures with errorbars. It is unclear to me outside of figures 4 and 7) what precisely the ranges on these errorbars represent. In a note on Table 3, the range values are described as propagated uncertainties of variables used in the covariance. Is that the case for the bounds of the errorbars in the remainder of the figures? Or are they representing

other statistics of the distribution (e.g. standard deviation like Figs. 4, 7)? More thorough descriptions of these errorbars in their respective figure labels are necessary for reader understanding. Additionally, I think some more detail on the sources of uncertainty (as mentioned in Table 3) would be helpful for overall understanding. Specific uncertainty values are mentioned once in the paper on L139 for MODIS-retrieved LWP. It is clear, however, that other sources of uncertainty – from satellite observations and from simulations – are considered for calculating the bounds in Table 3 (and perhaps elsewhere). To the authors' credit, the uncertainties inherent to the utilized satellite retrievals (particularly with respect to their analysis of the trough regime) are discussed in the paper. But to contextualize the results, more thoroughly and specifically describing the sources of uncertainty and how they've factored into the analysis is necessary.

Thank you for pointing this out. We would like to clarify that we use two different notions of “error” in the manuscript:

1. Error bars in Figs. 4, 7, 8

These are standard deviations (SD) from the binned statistics shown in each panel (i.e., spread of the data that fall in that bin), and they do not represent propagated analytical uncertainty.

2. “±” values in Table 3 (now Table 4 in revised manuscript) and for slope estimates

These are standard errors (SE) of fitted slopes for the $\partial \ln(LWP)/\partial \ln(N_d)$ and $\partial \ln(N_d)/\partial \ln(\sigma_{MBL})$.

As for the

$$\frac{\partial \ln(LWP)}{\partial \ln(N_d)} * \frac{\partial \ln(N_d)}{\partial \ln(\sigma_{MBL})},$$

The uncertainty SE is obtained by propagating the two slope SEs. Assuming independence,

$$\frac{Var(AB)}{(AB)^2} \approx \frac{SE(A)^2}{A^2} + \frac{SE(B)^2}{B^2}, \text{ hence, } SE(AB) \approx |AB| \sqrt{\left(\frac{SE(A)}{A}\right)^2 + \left(\frac{SE(B)}{B}\right)^2}$$

Here $A = \partial \ln(LWP)/\partial \ln(N_d)$ and $B = \partial \ln(N_d)/\partial \ln(\sigma_{MBL})$. We have revised the corresponding figure and table captions to state this.

As for the satellite instrumental uncertainties, we have added the following discussion in Section 2:

‘Although the relative errors in N_d retrieval can be significant at the pixel scale (Grosvenor et al., 2018), 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). 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.’

Furthermore, we assessed how satellite retrieval uncertainty could affect the LWP- N_d relationship $\mathcal{L}_0$:

Since satellite (MODIS) N_d study report the report normalized RMS differences of ~30–50 % against aircraft data, we modeled N_d errors as multiplicative, mean-one lognormal noise with coefficient of variation (CV) 30–50 %. In log-log Ordinary Least Squares (OLS) fits, this is equivalent to additive noise in $\ln N_d$. With log-space error standard deviation is $\sigma_\epsilon = \sqrt{\ln(1 + CV^2)}$ (≈ 0.29 for 30 % and ≈ 0.47 for 50 %), and $\mu = -\frac{1}{2}\sigma_\epsilon^2$, the effect in log-space is:

$$\ln N_{d,obs} = \ln N_{d,true} + \eta, \quad \eta \sim N(\mu, \sigma_\epsilon).$$

Hence, the OLS with noise in the predictor $\ln N_d$ would experience the regression-dilution effect and the expected slope would be attenuated as:

$$\mathbb{E}[\mathcal{L}_0] \approx \mathcal{L}_0 * \frac{Var(\ln N_d)}{Var(\ln N_d) + Var(\eta)},$$

which implies ~10–30 % potential slope damping. Consistent with this expectation, a Monte-Carlo sensitivity test, multiplying the satellite N_d with lognormal noise (coefficient of variation ~30-50%) and refitting for 1000 times, attenuated the bulk satellite slope from -0.192 (baseline) to a median of -0.157 (~18 % attenuation), with a 95 % sensitivity band of $[-0.170, -0.143]$. However, this

does not alter the sign or the comparative result we report, but it does indicate that observed magnitudes should be interpreted with caution.

For E3SMv2 N_d there is no instrument error, its uncertainty manifests as model bias and state-dependent spread. We therefore treat model N_d as as-simulated and discuss the biases in the results, rather than inserting an additional error term into slope SEs.

We have added the above discussion in the revised Section 3:

‘To gauge how satellite retrieval uncertainty affects the LWP versus N_d relationship, we note that satellite N_d studies report normalized root-mean-square differences of about 30 to 50% relative to aircraft data as in Gryspeerd et al. (2022). In log space a multiplicative N_d error is additive in $\ln N_d$, implying an expected slope damping of roughly 10 to 30%. Consistent with this expectation, a Monte Carlo test that multiplied satellite N_d by lognormal noise with coefficient of variation ~30-50% and refit the slope 1000 times reduced the bulk satellite slope from -0.192 to a median of -0.157, an attenuation of about 18%, with a 95 percent sensitivity band of -0.170 to -0.143. However, this does not alter the sign or the comparative result reported here, but it does indicate that observed magnitudes should be interpreted with caution.’

Specific comments:

L185: What ERA5 outputs are used for nudging the model, specifically? What is the relaxation time? More information as to how the model has been nudged is necessary.

We have added the nudging detail as below:

‘In this study, EAMv2 was run at standard resolution (~110 km) with the meteorology nudged to ERA5. The model was nudged toward the ERA5 zonal (U) and meridional (V) wind and temperature fields using a relaxation time of 6 h. This nudging reduces errors in the simulated meteorology, facilitating the examination of aerosol and cloud properties.’

L189: You define σ_{EXT} already on L129

Thanks, it has been corrected now.

References: Mechem et al., 2018, is cited throughout the paper, but seems to be absent from your list of references.

This reference is corrected listed now.

Table S1: I found myself frequently referencing Table S1 during my review of this paper and feel that it would be a useful inclusion in the main manuscript.

We have moved the Table S1 to main text.