

Review of ‘Investigating the cloud liquid water and droplet concentration relationship across cloud morphologies with machine learning’ (egusphere-2026-4426)

The manuscript applies the self-supervised image embedding of Geiss et al. (2024) to ~10 years of Terra MODIS imagery around the ARM Eastern North Atlantic site, clusters the embeddings and merges the clusters into 12 morphological regimes, and examines the relationship between liquid water path (L) and droplet number concentration (N_d) within and across those regimes. The authors conclude that the inverted-V persists within most regimes and at fixed morphology, with slopes that differ from regime to regime. The manuscript is written with clear logic and a coherent narrative flow. I would also like to applaud the authors for releasing the full processing and plotting code together with the manuscript. That is the kind of open practice the community should expect.

My main concern is that the manuscript shows that the shape of the inverted-V survives conditioning on cloud morphology, but has not yet shown how much of the relationship survives, or whether what survives is a cloud response or simply co-variability with meteorology and with retrieval artifacts. Settling that would turn a descriptive result into a quantitative one, and would also give the Discussion the substance that a dataset of this size deserves. In several places the physical mechanism discussed behind a result is either not offered or does not fit the mechanism invoked elsewhere in the paper. The comments below are offered with that in mind.

Main Comments

1. Section 2.2 keeps pixels with $\tau_c > 1$ and $r_e > 3 \mu\text{m}$, with no upper bound on r_e . The Collection 6 liquid retrieval is capped at $30 \mu\text{m}$ (Platnick et al., 2017), and the thin, partly cloudy pixels that $\tau_c > 1$ lets through are the ones where the retrieved r_e is most biased high (Zhang and Platnick, 2011; Painemal et al., 2025). As far as I can tell from *extract_chip_Nd_LWP_mean.py*, the field read is ‘Cloud_Effective_Radius’, which in the MOD06 product is the $2.1 \mu\text{m}$ retrieval, so the default and most heterogeneity-sensitive radius appears to be the one used. For comparison, the studies the filters are said to follow (L144) excluded $\tau_c < 4$ (Gryspeerdt et al., 2019) or treated $\tau_c \leq 3$ as optically thin (Goren et al., 2025), and the ‘appropriate filtering’ of Painemal et al. (2025) cited at L167 removes pixels with $\tau_c < 3$.

I wonder whether this is what shapes the low- N_d end of several figures. In Fig. 5a the conditional PDF has a hard left edge at $\ln N_d \approx 2$ (about 7 cm^{-3}) with the densest cells of the whole panel sitting on it, and in Fig. 5b the N_d -only network leaves a residual of -0.5 to -0.9 there. That edge is where the r_e cap would put it. Eq. 1 gives $N_d \propto \tau_c^{1/2} r_e^{-5/2}$, and for an adiabatic cloud $L = (5/9) \rho_w \tau_c r_e$, so for pixels pinned at $r_e = 30 \mu\text{m}$ one has $\tau_c \propto N_d^2$ and therefore $L \propto N_d^2$, i.e. $\ln L = 2 \ln N_d + \text{constant}$. A quick calculation with Eq. 1 at $r_e = 30 \mu\text{m}$ and $\tau_c \approx 10$ gives $N_d \approx 7 \text{ cm}^{-3}$, which is where the edge sits. The low- N_d mode of the **Subtropical stratocumulus** regime (Fig. 4e, $L \approx 15\text{-}30 \text{ g m}^{-2}$, $N_d < 10 \text{ cm}^{-3}$) likewise needs τ_c of about 1-2 and r_e of $20\text{-}30 \mu\text{m}$, i.e. the same thin, large- r_e pixels in which the retrieval is least reliable. The text argues thin, non-precipitating cloud as the explanation (L370), and that may well be right, but with what is shown it cannot be told apart from a retrieval boundary. The ascending-leg slopes in Fig. 4 are the quantities most exposed to this, and from *plot_cluster_LWP_Nd.py*, the fit uses every sample below the peak bin, i.e. from the smallest N_d present (about 6 cm^{-3}) up to roughly 20 cm^{-3} .

Could the authors report the fraction of valid pixels at $r_e \geq 29 \mu\text{m}$ as a function of N_d and of regime, show the τ_c and r_e distributions of the low- N_d mode in Fig. 4e, and repeat the ascending-leg fits and Figs. 5-6 with $\tau_c > 3$ (and, if the data allow, with the $3.7 \mu\text{m}$ r_e)? If the low- N_d leg is unchanged under the stricter filter, that would be a useful result in itself and would give the statement at L167 the support it needs.

2. The explanations given in Sect. 4.2 for the regime-to-regime differences pull in different directions, and I had trouble assembling one picture from them. The introduction (L43) attributes the ascending leg to precipitation suppression, which needs precipitating cloud at low N_d . At L362 the positive slope of **Closed cells** at $N_d < 30 \text{ cm}^{-3}$ is explained by the regime being ‘not precipitating or only lightly precipitating’, which is the opposite requirement, and later on at L371 the absence of a positive leg in **Subtropical stratocumulus** is explained by those clouds being non-precipitating. Across the panels of Fig. 4 the ascending slopes do not seem to follow precipitation tendencies either: the **Flowers** regime, whose defining feature is a precipitating cumulus core (Stevens et al., 2020), gives 0.04; **Sugar**, mostly cumulus humilis, gives 0.58; **Clustered cumulus** gives 0.69; **Storm-track stratocumulus**, which drizzles, gives 0.03. The steepest ascending legs belong to the most broken regimes, which are also the ones the authors flag at L543 as having the largest partly cloudy fraction. And also, the differences in the descending slopes (-0.12 for **Thick stratus** and -0.2 for **Open cells** against -0.44 for **Storm-track stratocumulus** and for **Mixed clustered cumulus and stratiform**) are noted at L377 but not explained.

One thing I noticed when looking at the supplement to the cluster mapping in *map_to_expert_labels.py* is how heterogeneous the regimes are in L - N_d space. Taking the means printed at the top of the supplement pages, the **Sugar** regime combines cluster 23 ($N_d \approx 118 \text{ cm}^{-3}$, $L \approx 43 \text{ g m}^{-2}$) with cluster 36 ($N_d \approx 28 \text{ cm}^{-3}$, $L \approx 121 \text{ g m}^{-2}$), and the **Closed cell** regime spans sub-clusters with mean L from about 58 to 122 g m^{-2} . So part of the within-regime inverted-V in Fig. 4 may itself be composition across sub-clusters.

Could the authors elaborate on this, and on what physically sets the ascending slope of a regime, and maybe order the regimes by ascending slope against a precipitation proxy (i.e., the fraction of pixels with $r_e > 15 \mu\text{m}$ as in Goren et al., 2025) and against partly cloudy fraction? From there, either a consistent mechanistic account or a plain statement that the within-regime ascending legs cannot be attributed to precipitation suppression from these data would be fine.

And I suggest a compact figure of the slopes of all 64 clusters, grouped by regime in the supplement, would also let the reader judge how much of Fig. 4 is sub-cluster mixing.

3. Fig. 5 holds what I think is the most important quantitative result in the paper, but it was discussed qualitatively. Eyeballing the red lines, the one in panel a) (the conditional median, per *plot_nn_residuals.py*) rises by roughly 1.0 in $\ln L$ between $\ln N_d \approx 1.9$ and 2.9 and falls by roughly 0.6 between 2.9 and 5.6 . The one in panel c), the residual after the morphology-only network, rises by perhaps 0.2 and falls by 0.25 over the same ranges. The difference between the two curves is exactly $E[f(\tilde{x}) | N_d]$, the part of the inverted-V that morphology alone predicts, and by my reading it is something like 80 % of the ascending amplitude and 60 % of the descending amplitude. Starting the ascending leg at $\ln N_d \approx 2.3$ instead (to stay clear of the retrieval edge I aforementioned in the first comment) the rise in panel a) is still several times that in panel c), so the proportion does not hinge on where the leg is taken to

begin. The shape survives conditioning, as stated in L485, but most of the amplitude does not, which is not exactly what ‘arises largely independently of cloud type’ (L511) or No. 4 of the Discussion (L531).

And yet, the variance-explained discussion in Sect. 4.3 points the other way. The authors further state that adding $\ln N_d$ to $\tilde{x}$ raises the explained variance by 4.21 % against 4.33 % for $\ln N_d$ alone (L467), which is stated as ‘largely independent’. But the inverted-V presented here is a feature of the conditional mean, and a conditional-mean feature can be mostly accounted for by a covariate while contributing almost nothing to R^2 either way. Both metrics are informative, and I would ask that both be reported, with the amplitude given at least equal weight.

Could the authors plot $E[f(\tilde{x}) | N_d]$ explicitly from panels a) and c), report the ascending and descending amplitudes and slopes before and after conditioning (with the spread across the eight trained networks)? The same decomposition can be done discretely with the 64 clusters, $E[\ln L | N_d] = \sum_k P(k | N_d) E[\ln L | N_d, k]$, with $\sum_k P(k | N_d) E[\ln L | k]$ as the composition-only curve. That would put a number on the question asked at L304, which Fig. 3 can only answer qualitatively, as the authors note at L341. It would also help to say why morphology accounts for so much of the amplitude. Presumably the regimes that occupy the low- N_d , high- L end (**Open cells** and the cumuliform regimes) are the deeper, precipitating boundary layers and those at the high- N_d , low- L end (**Subtropical stratocumulus**, **Closed cells**) the shallower ones, in which case the open question is whether N_d drives the morphology or the boundary layer drives both.

4. What holding $\tilde{x}$ fixed removes is an interesting point about Fig. 6, while the discussion seems not to address it adequately. For instance, aerosol perturbations can change morphology: the closure of open cells under ship emissions documented by Christensen et al. (2024) is a change in $\tilde{x}$, as are stratocumulus-to-cumulus transitions. The total derivative is the direct term in Fig. 6 plus a term that runs through morphology, and the Discussion around L568 recommends the method for chain-rule expansions of the aerosol forcing without saying that the second term has been dropped. I would ask that this be stated plainly and if feasible, that the mediated term be estimated. A regression of the $\tilde{x}$ components on $\ln N_d$ within regimes would give a first look. If the authors believe the morphology response to N_d is small in this dataset, that is itself a result worth stating with the reasoning.

Fig. 6 is also described as ‘how much the liquid water path is expected to change in response to changing N_d when cloud type is not allowed to change’ (L502), and the streamlines is noted are ‘how we would expect clouds to move in this phase space if N_d were increased’ (L509). What the figure shows is how L and N_d co-vary across scenes of the same morphology, i.e. a slope of the fitted surface. Meteorology is not held fixed, and the authors acknowledge this at L394. E.g., even within one morphology, the high- N_d scenes will tend to be those in shallower boundary layers nearer the coast, and the low- N_d scenes those in deeper boundary layers offshore. Deeper boundary layers carry thicker clouds, so a negative slope at fixed morphology can arise from this co-variability alone, with no cloud response to N_d involved. Could the authors clarify how far they expect this to affect Fig. 6, and maybe reword the figure as an association at fixed morphology? A quick check worth doing is whether the derivative changes when location is added to the $N_d + \tilde{x}$ network.

I am also curious about a physical interpretation I see in the figure, and I would like to hear the authors’ view on it. By eye, the sign change of the derivative moves from N_d of roughly 15 cm^{-3} at $L = 30 \text{ g m}^{-2}$ to

roughly $25\text{--}30\text{ cm}^{-3}$ at $L = 300\text{ g m}^{-2}$, and I wonder what sets its position and its drift with L . My guess is the onset of precipitation at a threshold r_e . If so, the sign change should follow a line of constant r_e , along which Eq. 1 with $L \propto \tau_c r_e$ gives $N_d \propto L^{1/2}$, so it should shift by about a factor of 3 over that range of L , against roughly a factor of 2 in the figure. Maybe the authors could test it by overlaying constant- r_e lines onto it?

5. The Discussion (Sect. 5) restates the four results and then turns to future work, and I came away unsure what a reader who models aerosol-cloud adjustments should take from the paper. Which slope is the relevant one for an effective radiative forcing estimate: the aggregate inverted-V, the within-regime slopes of Fig. 4, or the morphology-controlled derivative of Fig. 6? The three differ by a factor of several on the ascending leg, and the paper does not say which it recommends or why. Related to that, the manuscript never states in the abstract or the conclusions that its relationship is a 128-km scene-mean one, and Gryspeerdt et al. (2019) note that the slope changes with aggregation scale. Could the authors further clarify those?

Minor Comments

1. L105 and L120. Geiss et al. (2024) trained the CNN on the center 1024 of the 2708 half-kilometer pixels, i.e. viewing zenith angles below about 21° , whereas here the full swath is encoded for clustering (the 55° limit applies to the L2 pixels only, L151). Pixel growth and bow-tie distortion change the apparent texture scale toward the swath edge. Does cluster membership depend on viewing angle? A viewing-zenith histogram per regime would settle it. The ‘128x128 km’ in the Fig. 3 caption also holds only near nadir.

2. The adiabatic formula (L140) fixes $f_{ad} = 1$, $c_w = 2 \times 10^{-6}\text{ kg m}^{-4}$ and $k = 1$ for all regimes. Grosvenor et al. (2018) summarize ground-based estimates of f_{ad} at about 0.66 ± 0.22 , with lower values in cumuliiform cloud. Since $N_d \propto f_{ad}^{1/2}$, assuming $f_{ad} = 1$ overstates N_d in the cumulus regimes relative to the stratiform ones. Furthermore, the cloud-top temperatures printed in the supplement run from about 277 K (**Open cells, Storm-track stratocumulus**) to 285 K (**Subtropical stratocumulus, Clustered cumulus**), over which c_w is not constant (Grosvenor et al., 2018). And the aircraft value $k = 0.80 \pm 0.13$ summarized by Grosvenor et al. (2018) would rescale N_d by 25 % relative to $k = 1$, which moves the $20\text{--}30\text{ cm}^{-3}$ N_d peak in the manuscript. Though they won’t change within-regime slopes, it does move the regimes relative to one another in Fig. 3 and moves the peaks between regimes, so a sentence on regime-dependent adiabaticity and a temperature-dependent c_w (or a sensitivity line) would be welcome.

3. Section 2.2 does not say which effective radius (2.1 or $3.7\text{ }\mu\text{m}$) or which L (the MOD06 ‘Cloud_Water_Path’ product or a recomputed one) is used. I had to go to `extract_chip_Nd_LWP_mean.py` to find out. Please state both. Note that product L assumes one vertical profile while Eq. 1 assumes an adiabatic one. Though the mismatch is a constant factor and does not affect slopes, it should be acknowledged.

4. L170: chip means are computed over valid pixels only, with a 10 % valid-pixel threshold. In the broken regimes the mean may rest on 10-30 % of the pixels, selected toward thick cores by the τ_c and SPI filters. Could the authors show the distribution of valid-pixel fraction by regime and check whether valid fraction correlates with N_d within regimes?

5. L231 and the supplement: There is no table mapping the 64 clusters to the 12 regimes, so the 64 supplement pages cannot be connected to the main text. I could only do so from *map_to_expert_labels.py*, which also shows two discarded clusters (28 and 35), while the Sect. 3.2 percentages still add up to 100 over the remaining 62. A single table with cluster number, regime, sample count, mean L, N_d , cloud-top temperature and pressure, and the two slopes would sit better in the supplement, and the two discarded clusters should be identified too.

6. Fig. S35 (cluster 35) seems affected by sea ice, with scenes over the Labrador Sea in March-May (that may be why it is discarded), yet it carries mean N_d 124 cm^{-3} and L 88 g m^{-2} ‘for cloudy pixels’, so pixels in these scenes were passed by the cloud mask and by every L2 filter in Sect. 2.2 and received liquid-cloud retrievals. Was the surface or snow/ice flag considered?

7. L350 say Fig. 4 uses ‘the 25% of cases most similar’ to the cluster center. In *plot_cluster_LWP_Nd.py* the threshold is the 25th percentile of cosine similarity and samples above it are kept, which is the closest 75 %, whereas *plot_clust_mean_scatter.py* (Fig. 3b) keeps those above the 75th percentile (the closest 25 %) and *plot_cluster_sample_images.py* (Fig. 2) the closest 10 %. Which does Fig. 4 show? If it is 75 %, the text and the supplement caption need changing, and a sentence on how the slopes depend on the threshold would be worth adding either way.

8. L432-441: the hidden-layer widths are not given: *train_nn_models.py* uses 512 (256 for the 1024-vector input), 128 and 32 units, batch size 64, early-stopping patience 4 starting at epoch 3, against ‘minimum of 10’ epochs and ‘3 consecutive epochs’ at L437, and the parameter counts stated at L440 (1.4×10^5 and 8.4×10^3) do not look consistent with that architecture: a first layer of 512 units fed by 33 inputs already carries about 1.7×10^4 weights, more than the 8.4×10^3 stated for the whole model. In N_d is centered but not scaled to unit variance, contrary to L424. Please reconcile the text with the code, and state that Table 1 is evaluated on the 2021-2022 test years.

9. L483: ‘the inverted-V pattern is gone’ is not quite what Fig. 5b shows; the residual drops to between -0.5 and -0.9 at the lowest N_d bins and jumps to $+0.2$ just above, which is the edge discussed in the first main comment. The caption also says the red line is the mean, while *plot_nn_residuals.py* uses the median (*mean_yvals.append(np.median(lwp[(nd > x_bins[i]) & (nd <= x_bins[i+1])]))*)).

10. L505 and Fig. 6: Please state that the figure uses the 2021-2022 test years and the first fold's network, as *plot_model_gradients.py* does. The dots mark a within-bin t-test on sample derivatives. However, with thousands of samples per bin it is significant nearly everywhere, and it says nothing about model uncertainty. The chips also overlap by 50 % in each direction (stride 128 on 256-pixel chips, L120), so the effective sample size is roughly a quarter of the nominal one. I would suggest the authors fully acknowledge the limitation, or maybe report the spread across the eight trained networks.

Technical Comments

1. L20 still carries the placeholder copyright 'TEXT'.
2. L154: an effective-radius threshold of 3 μm has no effect, since the Collection 6 liquid-phase retrieval is bounded to 4-30 μm (Platnick et al., 2017).
3. Fig. 3 caption: ' $P(N_d|L)$ ' should be ' $P(L|N_d)$ ', as in the Fig. 4 caption and at L39?
4. Fig. 4 panel l) is 'Gravel and Sugar' while the text calls the regime 'Mixed sugar and clustered cumulus'.

References

- Christensen, M. W., Wu, P., Varble, A. C., Xiao, H., and Fast, J. D.: Aerosol-induced closure of marine cloud cells: enhanced effects in the presence of precipitation, *Atmos. Chem. Phys.*, 24, 6455-6476, <https://doi.org/10.5194/acp-24-6455-2024>, 2024.
- Geiss, A., Christensen, M. W., Varble, A. C., Yuan, T., and Song, H.: Self-supervised cloud classification, *Artif. Intell. Earth Syst.*, 3, e230036, <https://doi.org/10.1175/AIES-D-23-0036.1>, 2024.
- Goren, T., Choudhury, G., Kretzschmar, J., and McCoy, I.: Co-variability drives the inverted-V sensitivity between liquid water path and droplet concentrations, *Atmos. Chem. Phys.*, 25, 3413-3423, <https://doi.org/10.5194/acp-25-3413-2025>, 2025.
- Grosvenor, D. P., Sourdeval, O., Zuidema, P., Ackerman, A., Alexandrov, M. D., Bennartz, R., Boers, R., Cairns, B., Chiu, J. C., Christensen, M., Deneke, H., Diamond, M., Feingold, G., Fridlind, A., Hünerbein, A., Knist, C., Kollias, P., Marshak, A., McCoy, D., Merk, D., Painemal, D., Rausch, J., Rosenfeld, D., Russchenberg, H., Seifert, P., Sinclair, K., Stier, P., van Diedenhoven, B., Wendisch, M., Werner, F., Wood, R., Zhang, Z., and Quaas, J.: Remote sensing of droplet number concentration in warm clouds: A review of the current state of knowledge and perspectives, *Rev. Geophys.*, 56, 409-453, <https://doi.org/10.1029/2017RG000593>, 2018.
- Gryspeerd, E., Goren, T., Sourdeval, O., Quaas, J., Mülmenstädt, J., Dipu, S., Unglaub, C., Gettelman, A., and Christensen, M.: Constraining the aerosol influence on cloud liquid water path, *Atmos. Chem. Phys.*, 19, 5331-5347, <https://doi.org/10.5194/acp-19-5331-2019>, 2019.
- Painemal, D., Smith, W. L. Jr., Gupta, S., Moore, R., Cairns, B., McFarquhar, G. M., and O'Brien, J.: Can we rely on satellite visible/infrared microphysical retrievals of boundary layer clouds in partially cloudy scenes? Implications for climate research, *Geophys. Res. Lett.*, 52, e2024GL113825, <https://doi.org/10.1029/2024GL113825>, 2025.
- Platnick, S., Meyer, K. G., King, M. D., Wind, G., Amarasinghe, N., Marchant, B., Arnold, G. T., Zhang, Z., Hubanks, P. A., Holz, R. E., Yang, P., Ridgway, W. L., and Riedi, J.: The MODIS cloud optical and microphysical products: Collection 6 updates and examples from Terra and Aqua, *IEEE Trans. Geosci. Remote Sens.*, 55, 502-525, <https://doi.org/10.1109/TGRS.2016.2610522>, 2017.
- Stevens, B., Bony, S., Brogniez, H., Hentgen, L., Hohenegger, C., Kiemle, C., L'Ecuyer, T. S., Naumann, A. K., Schulz, H., Siebesma, P. A., Vial, J., Winker, D. M., and Zuidema, P.: Sugar, gravel, fish and flowers: Mesoscale cloud patterns in the trade winds, *Q. J. Roy. Meteor. Soc.*, 146, 141-152, <https://doi.org/10.1002/qj.3662>, 2020.
- Zhang, Z. and Platnick, S.: An assessment of differences between cloud effective particle radius retrievals for marine water clouds from three MODIS spectral bands, *J. Geophys. Res.*, 116, D20215, <https://doi.org/10.1029/2011JD016216>, 2011.