Authors' answers to Anonymous Referee #2

We thank the Anonymous Referee #2 for their valuable feedback on our manuscript, which improved the quality of the manuscript significantly. We answer their comments below, using blue font. Where necessary, we indicate the line numbers (Ln) of edits made to the manuscript like so: (Ln in original manuscript/Ln in updated manuscript).

Also please note that we found a tiny mistake in the standardization of the data and re-ran the analyses but the great majority of results remain unchanged: only one arrow in the graph goes from insignificant to significant, but it does not change sign. We added a sentence describing this arrow in the text.

Summary

The authors present a statistical analysis of the drivers of aerosol-cloud interactions in marine boundary layer clouds in satellite observations and in ICON Sappire, a cutting-edge global storm-resolving model. The statistical analysis aims to get at causal pathways driving changes in cloud properties and dynamics due to aerosol perturbations. As the aerosol/microphysics/radiation treatment and coupling in the model is somewhat simplified, not all of the same pathways exist in the two frameworks. However, the analysis yields some insight into the behavior of the model and its biases.

Recommendation: Minor revisions

This paper is advancing a new and seemingly promising statistical technique for understanding the (causal) mechanisms driving aerosol-cloud interactions. I am supportive of this paper being published in ACP, though after minor revisions. In the report below, I ask many questions about the causal approach and how the details or uncertainties in the satellite retrievals might affect comparisons with the model results. As I am not an expert in either statistical analyses related to causal analysis or satellite retrievals, I would ask the authors to think of them as questions driven by curiosity rather than as criticism of the manuscript and study. Perhaps, other readers new to these approaches might have the same kinds of questions. Some of them may not be worth addressing in the manuscript. Most all of the others might just require a reference or a sentence to address.

I would be willing to review the manuscript again if the editor thought it would be useful.

===========================

Major comments:

1. (p. 12/sec 3) I have some knowledge of cloud-aerosol interactions but little of causal approaches such as used in this study. I have certainly learned a lot about

them by reading this paper and some of its references. My questions below may be simple minded, but I might not be the only reader to have such questions. While addressing some of the questions might require a bit of extra exposition, I think only a sentence or two in the appropriate places would be sufficient.

Thanks to the comments from both reviewers, we have realized the description of causal methods was a bit terse. As explained in the Author's comments to Reviewer #1, we have streamlined the discussion of correlative vs. causal satellite studies in the introduction and added a few sentences at the beginning of section 3 to better introduce causal methods. We also added more details to the methods section.

 - My understanding from Fons et al (2023, specifically the caption to supplemental figure 12) is that the satellite causal graph reflects only daytime conditions. Is that true? If so, please state it clearly somewhere. (Apologies if I've missed that.) However, the availability of both day- and night-time data from ICON offers the possibility to see whether the causal networks change from day to night (at least in model world). That seems interesting, because marine boundary layer clouds have significant diurnal cycles of radiative cooling, entrainment, liquid water path and precipitation, which could (?) change the causal connections between day and night conditions.

That is a very good point, we forgot to add this discussion in the manuscript, thank you for the suggestion. The satellite retrievals for optical cloud properties only include daytime data points as they are based on radiance measurements in visible/near-IR channels. We have added a sentence mentioning this in the methods section (Ln 497/Ln 595).

The model results presented in the pre-print included both daytime and nighttime data. We additionally provide comparative daytime vs. nighttime results for the model in the updated version of the manuscript (Fig. 8, Ln 369/Ln 413-428). These comparative results nicely illustrate that adjustments will be different at night and during the day, implying a selection bias for daytime-only satellite studies.

We also now explicitly indicate in the figure captions whether the plotted reanalysis/model data correspond to daytime only or both daytime and nighttime, depending on the variable.

 - As GPM satellites with radar and/or microwave instruments are not continuously overhead, most precipitation data will rely on the same geostationary satellite that is also providing information about cloud optical depth, effective radius and cloud top temperature. When combined with the adiabatic assumption, there seem to be many fewer degrees of freedom in the satellite data than arrows in the causal diagram. How should the reader think about this?

We discussed this issue briefly in Fons et al. (2023). N_d, r_{eff}, H and LWP come from the same radiance measurements in the visible and near-infrared. Precipitation will be inferred from infrared measurements from the same satellite instrument, combined with polar-orbiting microwave measurements propagated along wind vectors. Arola et al. (2022) (https://doi.org/10.1038/s41467-022-34948-5) found that correlated retrieval errors in r_{eff}/COD satellite products, associated with the adiabatic assumptions, could result in spurious N_d-LWP associations. As you mentioned, this could also be true to some extent for precipitation measurements that come from the same satellite, despite coming from different wavelength channels. We have now added a sentence describing the impact of correlated noises in the conclusion (Ln 415/Ln 502).

 - Rainfall (measured at cloud base or the surface) matters to the dynamics of the boundary layer when it is strong and has little influence when it is weak (in comparison to the other drivers of turbulence and convection). If the impact of rainfall on the boundary layer is nonlinear, how will this show up in the causal relationships, which seem to be linear and based on the standardized changes of rainfall across all times?

It is true that the linear assumption could explain why the behavior of RR in the causal graph is not completely as expected. We have mentioned this in the discussion of the direct causal effects. If you could suggest a reference for the non-linearities you describe, we would be happy to add it to the discussion. For the moment, we have added a description of the precipitation arrows that will be impacted by the linear assumption (Ln 256/Ln 327): "*It should also be noted that the causal method used here is based on a linearity assumption which might be broken for non-linear precipitation processes. In particular, the lagged arrow from RR to N_d (wet scavenging and dynamical effects of precipitation) might be incorrectly captured in case of non-linear threshold effects of precipitation on the dynamics of the boundary layer. The onset of precipitation (arrow from r_{eff} to RR) is also non-linear but is expected to follow a power law, thus it should be well captured thanks to the log-transformation of the variables (see methods)*".

 - As susceptibility of low cloud quantities is often defined in terms of logarithmic changes (e.g., equation 4 in Terai et al, 2012, https://doi.org/10.5194/acp-12-4567-2012), would it be worth contrasting that briefly with the linear sensitivities represented by the alpha's and beta's in this study (although standardized as described as in Fons et al, 2003)?

As in Fons et al. (2023), the cloud variables are log-transformed in line with other studies in the literature, but we did forget to mention the log-transformation in the methods, thanks for catching the omission. This has now been added (Ln 520/Ln 619).

- Cloud fraction adjustments to increased aerosol occur in clouds with low background aerosol concentrations. Are they worth including in the causal network, especially if the pixel size for the causal analysis is 0.5 degrees?

The 0.5 ° average is performed on cloudy pixels only, so CF adjustments should not be captured by the data. We have also added this to the methods section (Ln 520/ Ln 617).

- Are the causal relationships in the network based on the high CCN or the low CCN simulation, or a combination of the two? My understanding of the time series analysis is that the causal relationships depend on time variations in predictor quantities in a single simulation. That seems quite different from other approaches that difference quantities between the high and low CCN simulations, so it would be good to be clear on this point.

That is a good distinction to make, we have added this to the methods section. The causal analyses are based on the 'low CCN' simulation, we have also added this to the methods section (Ln 538/Ln 645).

- (inspired by p. 17/fig 8) If the causal effects are computed from time series with an increment in predictor (e.g., aerosols) at time zero, wouldn't the influence of that local in time-and-space aerosol increase move away from a fixed Eulerian location as the winds advect the airmass? What is the meaning of these after-effects 24 hours later than an aerosol (or other) perturbation at a fixed Eulerian location? Does the aggregation of data to larger scales (here 0.5 degrees and 10 degrees in Fons et al, 2023) avoid this issue?

We discussed this in Fons et al. (2023) but added a brief discussion in this paper as well (Ln 331/Ln 373). The causal temporal developments are not calculated from an observed temporal evolution. Instead, they are calculated from the time propagation of the direct causal effects and autodependency coefficients (shown in Fig. 7). The direct causal effects are computed over 15-min intervals only, and we assume that there is minimal advection of cloud fields past the 0.5° boxes within this time frame. The resulting propagated temporal developments should be understood as hypothetical temporal developments should the clouds persist for this long, but they are not a direct measurement of cloud lifetimes.

- Also, do satellite retrievals of effective radius correspond to cloud top values or some type of integral over the depth of the cloud? Shang et al (2023, ACP, https://doi.org/10.5194/acp-23-2729-2023) suggest that different wavelengths may give information about different vertical levels.

r_{eff} satellite retrievals correspond to the effective radius close to cloud top. To be exact, this is only an approximation, but King et al., (1992) estimate that r_{eff} should correspond to the droplet radius at 85-95% of the cloud height for optically thick

clouds, i.e., close to cloud top. We added 'cloud effective radius *at cloud top'* to the methods to make this clear (Ln 496/Ln 593).

If answers to these questions are in references like Fons et al (2023), Runge et al (2019) or elsewhere, the authors could emphasize that in the text.

============================

Specific/minor comments (11/240 means p. 11, line 240):

 - Perhaps, the abstract could mention that the simulations are made for a period during boreal summertime.

This has now been added.

5/109: Did ICON try to include an estimate for subgrid vertical velocity variance in the parameterization for activation? While subject to uncertainty, that might have helped correct the low bias in Nd. However, this might not be worth such an effort when using a Smagorinsky-type parameterization, because they do not perform well in predicting subgrid quantities in simulations with these kinds of grid spacings. See e.g., Cheng et al (2010, https://doi.org/10.3894/JAMES.2010.2.3, sec. 3.1.2.1).

The activation parameterization used here does not include an estimate of sub-grid scale velocity variance. Propositions for how to include such an estimate to ICON-HAM km-scale set-ups have been discussed at the latest HAMMOZ workshop, but this is being the scope of our study.

8/Fig 4: Do metrics related to decoupling change between the low and high CCN simulations? The drizzling boundary layer might encourage such decoupling with rain evaporation below cloud base. The differing rain mixing ratios (and presumably) rain rates suggest different levels of precipitation-related subcloud evaporation in the two simulations.

We have evaluated boundary layer decoupling by looking at the height difference between the cloud base (CB) and the LCL, as suggested by Jones et al (2011) (doi:10.5194/acp-11-7143-2011). We find that most boundary layers are not decoupled, and that the high CCN simulation shows slightly increased decoupling compared to the low CCN simulation despite its lower rain rates. This might be due to the deeper boundary layer and higher entrainment rates in the high CCN simulation. However, we also note that, due to the relatively coarse resolution of the model around the inversion (100 + meters), it might be difficult to accurately estimate boundary layer decoupling as the height difference between the LCL and the cloud-base. Similarly, due to the coarse resolution, the inversion is not sharp enough, making it difficult to use the other decoupling metrics suggested by Jones et al. (2011). We have nonetheless added a supplementary Figure showing the CB-

LCL decoupling metric for the low vs. high CCN simulation for all four regions and we refer to it in the discussion of Fig. 4 (Ln 125/Ln 137-140).

In the outlook, we also added decoupling as an example of non-linear processes that could be mis-estimated with the linear causal approach (Ln 416 / Ln 504): *"Non-linear causal methods could be a good option to better estimate precipitation effects but also understand how potentially non-linear decoupling of the boundary layer can modulate the results presented here."*

Also, vertical velocity variance might be useful for thinking about activation and cloud formation, alongside the mean vertical velocity.

Fig. 2g gives an idea of the variance in cloud-base vertical velocity in the 0.5° x 0.5° averaged data. The variance in the original 5-km data is about 3 times the one in the coarser data. We have added a supplementary figure showing the same as Fig. 2g but using the original 5-km resolution data. Even in that figure, we see that the 75th percentile of cloud-base velocity is on the order of magnitude of a few 10s of mm.s^{-1}, which is quite low and confirms the argument that low updraft speeds are responsible for the low cloud droplet number concentrations.

9/Fig 5: The labels a/b in the caption refer to the wrong panels. Also, why not also show the same plot for the low CCN simulation, or at least a contour showing how the joint Nd-LWP distribution shifts from low to high CCN?

Thank you for catching the labeling mistake, we fixed it. We also added a cross-reference to Suppl. Fig. 2 where histograms of the low CCN simulation are provided as well. The reasoning for only showing the high CCN 2D histograms in the main text is that cloud droplet number concentrations are very low in the low CCN simulation, preventing a fair comparison to the satellite data histogram.

Is the joint distribution based on grid point/satellite retrieval footprint data? if one or both datasets coarsened to the 0.25 or 0.5 degree resolutions mentioned in the satellite data appendix, a short description of how the Nd or effective radius data are coarsened (e.g., average over all cloudy columns) would be useful.

We added the data resolution (0.5°) to the caption, and the methods now include a mention of the averaging method, which is in-cloud, not all-sky (Ln 517/Ln 617).

With the high CCN simulation having uniform "aerosol" concentrations, I would expect that changes in cloud base Nd in ICON are induced mainly by changes in the updraft strength, while satellite Nd changes are most likely dominated by actual aerosol variability. Is this consistent with the interpretation of the authors? If so, how should this contrast make us think about the results?

As alluded to on Ln 150/Ln169, the idea for this study is that we use the standard partial sensitivity decomposition for the LWP adjustment (e.g. Bellouin 2020). With these causal graphs, we focus on the $\partial LWP/\partial N_d$ term, but ignore the $\partial N_d/\partial$Aerosol term, which this comment is about. For this term, it does not matter why N_d changes, as long as it changes. We added this after discussing the term separation on Ln150/Ln 169.

"Separating the terms and focusing only on the first one also allows for a fairer comparison of the satellite and model data, as the model's $\partial N_d/\partial A$ term will not be defined due to the constant CCN assumption"

9/163: The "invisible ship tracks" paper of Manshausen et al (2022, https://doi.org/10.1038/s41586-022-05122-0) does show positive LWP adjustments in trade cumulus regions and might be worth mentioning here alongside Jiang et al.

Great suggestion. We have added the citation to Manshausen et al. (2022), but rather in the conclusion of the manuscript (Ln 405/Ln 485):

"In fact, Manshausen et al. (2022) used ship tracks to show that LWP adjustments are weakly negative in the stratocumulus cloud regimes, but positive in the trade cumulus cloud regime where most tracks are `invisible'. This is consistent with the cumuliform-looking ICON clouds showing positive LWP adjustments in our study.'"

10/Fig 6: Are the Nd and r_d values shown here averaged over all grid points (cloudy and clear) or only over cloudy grid points? The in-cloud values are a more useful reference when thinking about aerosol-cloud interactions. I would be surprised if the in-cloud values of N_d and r_d would go to zero so smoothly at the top and bottom of the cloud layer, but perhaps this does happen in ICON.

Fig. 6 shows all sky values, as they look nicer for plotting vertical profiles. We had explicitly mentioned this for the vertical profiles in Fig. 4 (see Ln 160-165). We had also included Suppl. Fig. 3, showing instantaneous cloudy-pixel-only values as histograms. We have now added instantaneous N_d and r_d values in Suppl. Fig. 3 and we now mention the all-sky average and Suppl. Fig. 3 when discussing Fig. 6 as well.

11/182 and 14/266: Does ICON Sapphire really use a fixed effective radius for all cloud liquid even when the double moment microphysics is used? I would understand this choice, but this should really be explicitly stated somewhere in this paper since I couldn't find it in the original Sapphire paper.

The radius can be diagnosed from the prognostic cloud liquid water content and the cloud droplet number concentration given assumptions about the size distribution. However, we never wanted to imply that r_d is constant. What we wanted to say here is that cloud droplet sedimentation is not parameterized because, due to the vertical extent of the grid boxes, cloud droplet sedimentation is expected to be a

negligible sink for cloud water compared to evaporation or rain autoconversion. We added this explanation of why cloud droplet sedimentation is turned off in the manuscript (Ln 181/Ln 207).

13/236: Could the magnitudes in the model graph be stronger because the model Nd is low and the clouds in the model simulation are precipitating more strongly than in the satellite data? Is it clear that the alpha's should be invariant as the background aerosol/Nd changes?

If the relationships are indeed linear, then the alphas should be invariant. However, if they are not in reality linear, the alphas could vary. We had discussed this linearity assumption in the analyses and in the discussion but not explicitly at this specific location of the manuscript. We added the following sentence after mentioning the different magnitudes in the model and in the satellite data: "Another explanation could be that the ICON graph captures a non-linear lower-N_d/higher-RR regime, potentially causing the alphas to vary in magnitude due to the linear assumption behind Wright's path analysis." (Ln 237/Ln 270)

top of p. 14: For the reader, it might be helpful to first talk about alpha's with clear signals in the satellite and model before later talking about those that are more complicated.

We moved the discussion of precipitation influences, which are less well quantified by the causal approach, to after the discussion of entrainment influences, which are more straightforward.

15/304: Stevens and Seifert (2008, https://doi.org/10.2151/jmsj.86A.143) is an earlier study along these lines. Albrecht (1993, https://doi.org/10.1029/93JD00027) isn't a direct predecessor but makes clear the impact of precipitation on the depth of the marine boundary layer.

Thank you for the suggestion, we have added the citation to Stevens and Seifert (2008). (Ln 304/Ln 344)

19/420: Possible reference for "increasing the vertical resolution": Bogenschutz et al (2023, https://doi.org/10.5194/gmd-16-335-2023). More sophisticated subgrid turbulence closures also offer some promise of better representation of stratocumulus at kilometer-scale grid spacing: Shi et al (2018, https://doi.org/10.1175/JAS-D-17-0162.1), Bogenschutz et al (2023, https://doi.org/10.1029/2022MS003466).

Thank you for the suggestions! We have added these references.

20/449: Specify a particular altitude, not one relative to the inversion height (which is variable).

We modified the sentence to be more descriptive of the actual vertical resolution (Ln 449/ Ln 540): *"The vertical resolution increases progressively from 20 m at the surface to 400 m at an altitude of 8 km, with vertical resolutions around 125 - 160 m at the inversion level."*

22: It would be worth mentioning the footprint of the satellite retrievals from SEVIRI, to give some context about how many pixels are being averaged into the 0.25 or 0.5 degree grids.

This has been added on Ln493/Ln588.

===========================

Typographical/rephrasing suggestions (OPTIONAL):

12/215: "along" --> "alongside"

16/318: "datam" --> "data"

19/421: "e.g.," should be before the reference. Maybe add another <> before the {PossnerEtAl2014}?

20/444: "a an" --> "an"
Citation: https://doi.org/10.5194/egusphere-2024-195-RC2

These typos have been corrected, thank you for catching them.