Authors' answers to Anonymous Referee #1

We thank the Anonymous Referee #1 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.

This is an interesting and well-presented analysis of aerosol-cloud adjustments in a convection-permitting model over Sc decks. The analysis technique (causal graphs) is novel and is used to compare to observations. The paper is well-written. I have a few concerns that are detailed below related to the appropriateness of comparing a convection-permitting model with fixed cloud condensation nuclei to observations using this particular framework. It is my opinion that these concerns can be handled by adding caveats.

If the authors find it appropriate and not excessively time consuming, they may find that the robustness of their analysis and their suggestion that there is a strong benefit to high resolution would be aided by including analysis of a lower resolution version of the same model with interactive aerosol and with fixed CCN so that they provide examples to argue (i) that their assumption of fixed CCN is not affecting their results and (ii) that the high resolution model actually improves model performance. In particular, constant CCN makes it a bit hard to interpret the causal graphs used in this analysis. To use such a high resolution model some parts of the simulation need to be sacrificed, but the ability of the aerosol, cloud, and precipitation to couple is key to reproducing observed present-day covariability and fairly compare observations and models (Stevens and Feingold 2009; McCoy et al. 2020; Wood et al. 2012; Gryspeerdt et al. 2019). Is there a way to modify the causal graph if one arrow linking precipitation, CCN, and cloud is disabled in the model, but exists in the observations?

We agree that it would be very interesting to conduct the analyses suggested here. However, the coupling of the ICON model with the HAM aerosol module is not finalized and extensive scientific validation still needs to be carried out. This is the topic of ongoing work at ETH Zürich and is the reason why we resorted to using an ICON version with constant CCN concentrations. Results from very few ICON-HAM GCM simulations (Salzman et al., 2021; https://doi. org/10.1029/2021MS002699) have been published in the literature, but these were based on an old version of the ICON model, and it would not make sense to use this model version today. It would be interesting to conduct a full sensitivity study with respect to resolution and to the consideration (or not) of the HAM module when ICON-HAM is ready, but this is beyond the scope of this paper. We now mention these considerations in the conclusion paragraph for clarity (Ln 402 of original manuscript/Ln 478-481 of revised manuscript).

Concerning how this affects our analyses: the only causal graph arrow that would be affected by the constant CCN assumption is the lag-1 arrow from RR to N_d, which is expected to describe both wet scavenging and (thermo-)dynamical boundary layer adjustments to precipitation. With a constant CCN assumption, this arrow would be left to describe only the second process. This is the reason why this lagged arrow is (weakly) negative in the observations but positive in the model. Of course, the absence of wet scavenging could explain why LWP adjustments are different in the model and in the observations. The point of applying the causal graph to the constant CCN model is to evaluate other hypotheses that might also explain this difference, namely: wrongly simulated entrainment enhancement and cloud deepening.

Throughout the paper, we streamlined the discussion a bit more to make it clearer that wet scavenging is another hypothesis that can explain the different observation-model LWP adjustments. For example, we added this to the abstract: *"We find that the positive LWP adjustment to increasing aerosols in ICON results from a superposition of processes, with an overestimated positive response due to (1) precipitation suppression, (2) a lack of wet scavenging, and (3) cloud deepening under a weak inversion, despite (4) small negative influences from cloud-top entrainment enhancement"*; and we added this to section 2.2 (Ln 174/Ln 195-199): *"Because CCN concentrations are kept constant, the causal link for wet scavenging is turned off. Wet scavenging has been shown to induce a negative correlation between N_d and LWP (McCoy et al., 2023), so turning it off biases the correlation towards more positive valuesThese two mechanisms alone (precipitation suppression and no wet scavenging) could explain the positive sign of LWP adjustments in the model. In the rest of this paper, we focus on other hypotheses for this observation-model discrepancy, namely: entrainment effects and cloud deepening."*, and we also mentioned the lack of an explicit parameterization for aerosol-induced entrainment enhancement and cloud deepening in our ICON version, justifying the utility of the causal graph for the study of these two mechanisms.

Ln5: Observing = inferring if it is based on correlation. If there is a transient change in aerosol it may be possible to say there is an observe a causal response. On ln 39 the cited studies are correlative, rather than causal.

Thanks to the reviewers' comments, we have realized that the discussion of causal inference methods was a bit terse. We added a few sentences to the introduction to distinguish the studies that are correlative vs. the studies that implement causal inference methods, and what some of those methods are (Ln 40/Ln 40-50):

"[…] On the contrary, satellite analyses show a negative Nd-LWP relationship, implying a dominance of cloud-top entrainment enhancement (e.g., Gryspeerdt et al., 2019; Possner et al., 2020). However, correlative satellite studies might overestimate this negative LWP adjustment by ignoring confounding meteorological influences. Several studies have used causal methods (Pearl, 2009) to remove spurious biases from aerosol-cloud interactions, either by directly targeting environmental confounders (Gryspeerdt et al., 2016; Varble, 2018), simultaneously removing multiple confounding influences through multivariate regressions (e.g, Andersen et al., 2017; Wall et al., 2022), or by using opportunistic experiments to indiscriminately remove all environmental confounding (Christensen et al.,

2022). For instance, Toll et al. (2017, 2019) and Chen et al. (2022, 2024) used ship tracks and/or volcanic eruptions to demonstrate that LWP adjustments of low-level marine clouds to aerosols are close to zero or slightly negative. Even when environmental confounding is removed from satellite studies of LWP adjustments, they still disagree with the strong positive adjustments seen in some GCM simulations. […]"

Ln 5: While high resolution models are probably generally good since fewer processes need to be resolved, we cannot provide a climate model that resolves the micro-scale processes that lead to precipitation suppression and entrainment thinning.

We changed the abstract sentence to reflect this comment: *"With higher-resolution global climate models, which allow the simulation of mesoscale circulations in which stratocumulus clouds are embedded, there is hope to start bridging this gap."* We do not add more detail in the abstract due to the word limit, but we do mention later in the introduction that cloud microphysics, turbulence and radiation still need to be parameterized.

Ln 35: Also (Wall et al. 2022).

Thank you for the suggestion. We added this citation to the list of studies that use causal methods for LWP adjustments estimations (see two comments above), as Wall et al. (2022) control for meteorology with the cloud-controlling factor approach.

Ln 40: While I generally agree that models probably overpredict adjustments due to minimal representation of size-dependent entrainment (Karset et al. 2020). Please note that the cited papers here are correlative rather causal. As discussed later in the intro (Christensen et al. 2022) there are opportunities for dealing with causal ambiguity in observations of aerosol-cloud interactions. Also note that in (McCoy et al. 2020) both models and observations imply a negative adjustment from aerosol-cloud interactions in the present day based on similar techniques to (Gryspeerdt et al. 2019) even though the model response to anthropogenic aerosols is an increase in LWP.

As mentioned above, we modified the introduction to clearly separate correlative studies from studies that implement causal inference principles.

Following your above comment on wet scavenging, we also added a discussion of the implications of the constant CCN assumption, and how this could explain the positive N_d-LWP relationship, and we now explicitly cite McCoy et al. (2020) and their discussion of the negative impacts of wet scavenging in section 2.2.

We have also added the citation to Karset et al., 2020 in section 2.2 (Ln 166/ Ln 186) when discussing size-dependent entrainment parameterizations, thank you for the suggestion.

Ln 44-59: I agree with all of this, but I think it would be good to qualitatively state what the range of length scales of interest are in Sc clouds. They are still definitely sub-km (Wood 2012).

We have added a sentence describing the spatial scales of interest for the study of stratocumulus clouds, from the sub-meter scale to the mesoscale (Ln 50/ Ln 60-62).

Ln 65: While fixed CCN is useful especially in high resolution models where running full aerosol chemistry would be expensive, it is somewhat unphysical since scavenging can't remove any CCN. This is an aspect of present day covariability between clouds, precipitation, and aerosol that seems like it would be needed when comparing between observations and the model output. Discussions of the source of covariability that goes cloud->precipitation->aerosol see discussion in (Gryspeerdt et al. 2019; McCoy et al. 2020; Wood et al. 2012). It would be interesting to contrast the analysis in this study with a lower resolution version of the same model with interactive aerosol and without interactive aerosol.

Unfortunately, as mentioned above, the ICON GSRM is not available yet with interactive aerosols. We do believe LWP adjustments in ICON need to be re-evaluated when ICON-HAM is ready, to compare how the adjustments operate with and witout HAM, and at different resolutions. This has now been added as an outlook to the discussion section.

However, the appeal of the causal graph is that one can separate physical processes. While it is evident that the causal arrow from precipitation to cloud droplet numbers will be affected by the constant CCN concentration, the other arrows should not, allowing us to separately investigate other processes that might differ for LWP adjustments in the observations vs. in the model, namely: entrainment and cloud deepening. As mentioned above, we streamlined the discussion in section 2.2. to make this reasoning clearer in the manuscript.

Ln70: One confounder is that aerosol sources tend to be on land, so there is spurious correlation between lower LWP and aerosol (Wood et al. 2012; Gryspeerdt et al. 2019). I am not sure nudging really impacts that all that much and since you have set CCN to a constant through the atmosphere this source of covariance is removed. Another strong source is that thicker clouds tend to rain more- reducing Nd and creating covariance that goes from clouds to precipitation to aerosol (McCoy et al. 2020; Wood et al. 2012; Gryspeerdt et al. 2019). This is also artificially turned off if CCN is set to a constant.

Ln 90: as the authors note- the simulations are not nudged, so it is hard to know how to compare these simulation by eye in the absence of comparisons of the ICON meteorology to observations/reanalysis.

We answer these two comments together:

We have reformulated the discussion on Ln70. First, as mentioned above, we moved the discussion of confounding that was previously on Ln70 up to Ln 40. Then, we made it clear that the point of the causal graph is to focus on physical processes rather than emerging statistical relationships, i.e. allowing a comparison in terms of process sign and strength between the observations and the models (Ln 70/Ln 81-87).

"It is challenging to do a direct comparison of aerosol-cloud interactions in the satellite data and in the model data as the simulations are not nudged to observations. As a consequence, as discussed in Sec. 2.2, the direct comparison of emerging statistical correlations between

aerosols and clouds can be tricky to interpret as they can result from different meteorological backgrounds. Instead, in Sec. 3, we use the causal methodology described in Fons et al. (2023) to disentangle superimposed processes occurring in aerosol-perturbed clouds, like precipitation suppression or entrainment enhancement. This focus on physical processes, rather than statistical associations, allows us to compare the response mechanisms of stratocumulus clouds in the model and in the observations, while removing some confounding originating from the effect of meteorology on entrainment or precipitation."

To be transparent, we also added in the conclusion some sources of meteorological confounding (relative-humidity, large-scale transport) that the present study neglects for the sake of simplicity (Ln 415/Ln 500). As you mention, the fact that aerosol sources tend to be on land is an example of neglected confounding. Concerning the fact that thicker clouds tend to rain more and scavenge more aerosols, this should be captured in the causal graph through the path $H \rightarrow r_{\mathrm{eff}} \rightarrow \mathrm{RR} \rightarrow N_{\mathrm{d}}$, with the exception that, in ICON, the RR –> N_{d} arrow only describes dynamical adjustments since wet scavenging is turned off. This is described in section 3.1. of the manuscript.

Ln 100: Agree- ERA5 is just running reanalysis atmospheric structure through a parameterization- albeit one where aerosol-cloud interactions are not being represented. It is good that the authors explicitly point out that ERA5 clouds are not reanalysis in the same way that things like near surface temperature are.

We agree, thanks for the comment.

Ln 108: Why isn't it also explainable by low CCN? Nd is a function of CCN and updraft and the constant CCN could be unrealistically low.

In the ICON (low CCN) simulation, CCN concentrations are actually not that low for marine regions in absolute terms (250 cm^{-3}), but they are low compared to the concentrations in ICON (high CCN), which are set to 1700 cm^{-3} . When comparing the CCN concentrations to the cloud droplet number concentrations (Fig 2b), one sees that only 5% of CCN are activated on average, which is very little. By checking the CCN activation parameterization, we conclude that this is due to low vertical velocities at cloud base.
We added these explanations to the methods (Ln 454/ Ln 547): *"For the sake of simplicity, these 2 simulations are referred to as `ICON (low CCN)' and `ICON (high CCN)' in this article, even though `low CCN' can be misleading since 250 cm^{-3} corresponds to typical CCN concentrations in stratocumulus-topped marine boundary layers (Roberts et al., 2010; Allen et al., 2011; Wang et al., 2022; Howes et al., 2023)*." These citations analyze campaign data and also show that activation rates are usually much larger than 5%.

Figure 4: Would be good to note that ERA5 RWC and LWC are the reanalysis thermodynamic fields run through a parameterization.

We have added a sentence concerning the parameterization of LWC and RWC in the discussion of Fig. 4 (Ln 131/Ln 149).

Line 171: Perhaps cite (Karset et al. 2020)

Thanks for the suggestion. This reference fits very well for the description of size-dependent entrainment effects and how that is taken into account (or not) by models, so we added it to the manuscript as mentioned above.

Figure 7 and Table 1: I am not familiar with the assumptions underlying causal graphs and it would help readers such as myself to understand and interpret these results if there was some discussion of how turning off arrows in the model like CCN being able to interact or size-dependent entrainment that we respectively know and suspect exist in the real world affects the causal graph. This seems like a broad discussion of comparing models, which will always be somewhat structurally incomplete to reality.

The causal arrows need to represent actual processes that occur (or are suspected to occur) in the considered data set. For this reason, it does not make sense to have an arrow from r_{eff} to w_{e} in this model because it does not parameterize size-dependent entrainment enhancement, even though this process does occur in reality, explaining why it is included in the satellite graph. Of course there is some level of subjectivity, as the person drawing the graph is deciding which arrows to include and which ones not to include. However, these choices need to be justified by physical arguments. We have added the following explanation at the beginning of section 3.1 (Ln 218/Ln 247):

"Fig. 7 shows the assumed N_d-LWP causal graph used in the analyses. A causal graph encodes domain knowledge by qualitatively describing causal relationships (arrows) between variables of interest (nodes). Each directed arrow describes a causal effect and not merely a statistical association, and can be justified by an underlying physical process. The colors of the arrows show the magnitude and sign of the direct causal effects $\alpha_{Xi,Xj,Iij}$ calculated from the causal graph."

The simplest way to understand what removing an arrow means statistically is by comparing to cloud-controlling factor analyses, such as the ones in Wall et al. (2022): by turning off the arrow pointing from a variable X to a variable Y, we are essentially saying that X does not need to be considered in the multivariate regression of Y with respect to its controlling factors. While multivariate regressions used in the aerosol-cloud literature (e.g. Wall et al., 2022 or Andersen et al., 2016) use varying cloud-controlling factors without necessarily justifying the choices other than saying that these variables are known to 'control' clouds, the causal graph allows to explicitly and very transparently justify why controlling factors are included. We added this short explanation to the methods section (Ln 532/Ln 634).

Concerning subjectivity of the arrow choices: Sensitivity studies can potentially be conducted to evaluate the effect of causal graph assumptions on the results. We did that to some extent in Fons et al. (2023). We would encourage other scientists in the field to come up with their own assumed causal graphs to be compared to ours. We believe this would generate interesting scientific discussions. We briefly discuss the dependence of the results on the assumed causal graph in the conclusion section.

Section 4: This section does a good job summarizing the results of this analysis.

Thank you!

References:

Christensen, M. W., and Coauthors, 2022: Opportunistic experiments to constrain aerosol effective radiative forcing. Atmos. Chem. Phys., 22, 641–674, https://doi.org/10.5194/acp-22-641-2022.

Gryspeerdt, E., and Coauthors, 2019: Constraining the aerosol influence on cloud liquid water path. Atmos. Chem. Phys., 19, 5331–5347, https://doi.org/10.5194/acp-19-5331-2019.

Karset, I. H. H., A. Gettelman, T. Storelvmo, K. Alterskjær, and T. K. Berntsen, 2020: Exploring impacts of size-dependent evaporation and entrainment in a global model. Journal of Geophysical Research: Atmospheres, 125, e2019JD031817.

McCoy, D. T., P. Field, H. Gordon, G. S. Elsaesser, and D. P. Grosvenor, 2020: Untangling causality in midlatitude aerosol–cloud adjustments. Atmos. Chem. Phys., 20, 4085–4103, https://doi.org/10.5194/acp-20-4085-2020.

Stevens, B., and G. Feingold, 2009: Untangling aerosol effects on clouds and precipitation in a buffered system. Nature, 461, 607–613.

Wall, C. J., J. R. Norris, A. Possner, D. T. McCoy, I. L. McCoy, and N. J. Lutsko, 2022: Assessing effective radiative forcing from aerosol–cloud interactions over the global ocean. Proceedings of the National Academy of Sciences, 119, e2210481119, https://doi.org/10.1073/pnas.2210481119.

Wood, R., 2012: Stratocumulus Clouds. Mon. Weather Rev., 140, 2373–2423, https://doi.org/10.1175/MWR-D-11-00121.1.

Wood, R., D. Leon, M. Lebsock, J. Snider, and A. D. Clarke, 2012: Precipitation driving of droplet concentration variability in marine low clouds. J Geophys Res-Atmos, 117, n/a-n/a, https://doi.org/10.1029/2012jd018305.