Overall comment

This study explores roles of spatial and temporal aggregation on quantifying albedo susceptibility (S_o) biases by analyzing the outputs of an ensemble of 127 large eddy simulations of marine stratocumulus. The authors designed three methodologies (L2, L3, and $L2_N$), which mimic common satellite-based analyses in different ways, to identify the influences of the adiabatic drop concentration Nd retrieval, the correlation between aerosol and cloud fields, and the extent of reduced variance in cloud albedo and Nd. The LES simulations also provide an opportunity to obtain the 'true' Nd, by which the effectiveness of adiabatic assumption used in satellite retrievals can be assessed. As a consequence of such an analysis, the authors obtained a series of interesting results regarding the influencing factors on albedo susceptibility biases. I think this is a very nice study, and the results presented also have significant implications for reconciling currently diverse observation-based estimates of aerosol indirect effects.

I would recommend this paper be published in Atmospheric Chemistry and Physics after my specific questions/concerns listed below are addressed appropriately.

Thank you for the careful review. Text additions are in blue and changes in red.

Specific comments

Line 106-111: If I understand correctly, B (in Eq. 9) is only relevant to the sensitivity of L to Nd, i.e., dln(L)/dln(Nd), not to the whole albedo sensitivity. It's a bit confusing for me how authors translate B to the theoretical calculation of the S_o biases?

Thanks for this important comment. The calculation was not explained. We revisit the methodology here:

1) We assumed for simplicity that $S_o \sim (1\text{-}A)/3$ (i.e., no L adjustments)
2) We assumed $A \sim \tau$ (reasonable at low enough τ)
3) Our two interacting fields are L and N
4) Therefore $\tau \sim$ L^(5/6) N^(1/3)
5) We calculated B_x and B_y (where x=L and y=N, β_L =5/6, β_N=1/3) and B_{cov}
6) Finally, the overall bias was calculated according to Eq. (9)

We do note that this is simplified, and representative of interacting L and N with their co-variabilities and spatial heterogeneities. When embarking on the LES-based calculations, we use τ and r_e as our key variables (to conform with the satellite approach), and L and N are derived from τ and r_e. In this sense our Fig. 1 calculations are distinct from the LES calculations and are therefore mostly illustrative. See text additions on lines 115-121.

Line 128: As for the LES, why did the authors only choose "nocturnal" instead of "diurnal" simulations or both?

We had performed a large set of simulations for a previous study and since these were readily available, we chose to use them. The important point is that they represent many different cloud scenes, ranging from closed- to open-cells, and a range of boundary layer/cloud depths, and cell sizes.

Line 139: How did the authors determine γ value in the calculation of Ac from simulated cloud optical depth?

As noted in the original manuscript on line 139 (now lines 157-158), A_c is calculated based on the modeled value of τ (Eq. 4).

Line 184: At what spatial resolution is cloud fraction defined here? Is it at 48km x 48km, or defined at 800m and 6 km respectively and then averaged up for whole domain?

Cloud fraction is defined at the native resolution since we're using it as a measure of the cloud field property. This is now clarified on line 204.

Line 186: It is expected that high fc (homogeneous clouds) would be associated with low bias in S. Why is the opposite here?

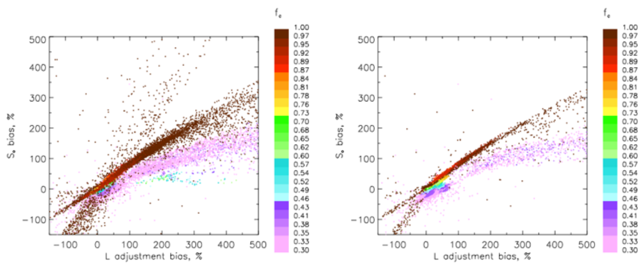
The way we chose to write this was as a process of discovery. One has to dig more deeply into the analysis before one can answer this question. But we have now added text to explain that the reason for this unexpected result will be revealed later (line 207). An easy, quick explanation is seen in Fig. 11. The adiabatic Nd retrieval in this high cloud fraction scene generates far too much heterogeneity, which results in strong S_o bias.

Line 184-192: The comparison between L2 and L3 methods here is to illustrate the aggregation biases associated with Jensen's inequality. Actually, there is already another practical method accounting for this issue based on satellite observations. For example, the MODIS L3 product includes a cloud optical depth-effective radius joint histogram which was suggested to consider the non-linearity in the calculation of Nd (e.g. Quaas et al., 2008; Grandey and Stier, 2010). Thus, it might be interesting to evaluate the effectiveness of this method from the LES data in this study.

Thank you for bringing these studies to our attention. The dependence of ACI metrics on scale in Grandey and Stier (2010) is relevant, although we note that they used level 3 data and aggregation scales ranging from 1-degree to 4-degree to 60-degree boxes. We now cite this paper in the revised text (line 80-83) but do not engage further because we don't see a straightforward way of connecting to our paper without major digression.

Line 274: How did the authors select these 58 simulations? Is there an objective criterion? Will it introduce artificial selection on cloud regimes? To show the robustness, it is useful to present the results from all 127 simulations, at least in the supplement information.

Since the 58 simulations are Latin Hypercube samples of the total, they are not biased. For example, the figure shows the full 127 simulations at left compared to the subsample at right. Because the points tend to align with a clear linear relationship, they quickly obscure points below, particularly at the intermediate cloud fractions. Text has been added on line 297-298.



Line 280: It's interesting that the separation of these branches for $L2_N$ is not as evident as L2 and L3. What is the underlying reason? The authors should explain in more detail.

The reason is that with the correct Nd, $L2_N$ avoids the bias and variance in Nd associated with clouds that have morphological structure. It is not only broken clouds but all clouds for which Nd retrievals generate unrealistic structure. This is now stated in the revised text on lines 306-307.

Line 300: Generally, a negative bias in retrieved Nd is expected duo to a positive bias in CER and a negative bias in COT for spatially inhomogeneous scenes according to the Eq.2 (Grosvenor et al., 2018). Thus, it is kind of surprising that the retrieved Nd for open-cellular clouds is larger than the true Nd.

Remember that we are assuming that r_e and τ are taken directly from the model, with no assumptions. In other words, we are not dealing with the broken cloud r_e and τ biases. We are only focusing on the effect of averaging in clouds with different morphologies. We now clarify this in the text on lines 333-338.

Line 317: It is not quite clear that how the authors conducted the regression fits. As for $\sum S_o$, does the 'individual scenes' here mean the whole domain? In this case, the regression fit was conducted over all 4x4- (or 30x30-) resolved grids in each scene, and then $\sum S_o$ was calculated by averaging up the individual S over all scenes (including the variations along both time and different simulations). Please clarify more detail on how the authors conducted the analysis.

The $\sum S_o$ is calculated as follows: first each scene (i.e., the entire domain) is used to calculate the sum; then the individual So values are averaged. This is done for all scenes that meet the criteria discussed in Section 3, and applied to all the other analyses. The second approach includes the same data, but now the regression is performed on all the aggregated data. In other words, the relationship now removes the natural co-variability associated with individual scenes. We now explain this more carefully in the revised text on lines 349-351.

Grandey, B. S., and P. Stier (2010), A critical look at spatial scale choices in satellite-based aerosol indirect effect studies, Atmos. Chem. Phys., 10(23), 11459–11470, doi:10.5194/acp-10-11459-2010.

Grosvenor, D. P. et al. Remote sensing of cloud droplet number concentration in warm clouds: a review of the current state of knowledge and perspectives. Rev. Geophys. 56, 409–453 (2018)

Quaas, J., O. Boucher, N. Bellouin, and S. Kinne (2008), Satellite-based estimate of the direct and indirect aerosol climate forcing, J. Geophys. Res., 113, 05204, doi:10.1029/2007JD008962.

Reply