Paulus et al. showcased Lagrangian LES simulations of Arctic marine cold air outbreaks based on the recent observational field campaign. Their major scientific objective is to understand how large-scale subsidence affects the development of clouds during the cold air outbreaks. The paper is clearly written and supported by elegant figures. The topic and results are of great interest to the community. I have a few questions regarding the modeling setup and conclusions. I would recommend publication if they are addressed.

Major comments:

- Regarding the modeling setup, there is strong nudging (3 hr time scale) imposed above the ABL. The buffer layer is very shallow (only 100 m), which makes me wonder if the evolution of clouds in the model is too heavily constrained by the nudging. Can the authors demonstrate the impact of this nudging on the cloud development by performing one or two sensitivity tests? This can help quantify the dependence of the results on the modeling setup.
- How do we reconcile the results from this study with the conclusions from Young et al. (2018)? Young et al. (2018) found that in their simulations, LWP and precipitation rates increase with subsidence rate, while IWP is not sensitive to subsidence changes. This is even true when surface warming is included. Here, LWP, IWP, and total precipitation all decrease with increasing subsidence. There is no sign of non-monotonic response to subsidence rates. The range of subsidence rates doesn't seem to explain this difference.
- While it may be beyond the scope of this paper to explore the sensitivity of the graupel formation timing to microphysics, it would be helpful to discuss how the microphysical scheme may affect the implications of the results.

L150–152: How well is ERA5 capturing the surface temperature evolution during this period? If surface parameters are from ERA5, it would be helpful to know if ERA5's surface temperature has any biases.

L208–211: I'd like to see more details about the bottom boundary conditions. Is the surface temperature prescribed using the observed surface skin temperature? How is atmospheric specific humidity calculated at the marginal ice zone? What is the temporal resolution of these bottom boundary conditions? I assume they are time varying.

L335: Remove parentheses around reference. Why is there a cold bias?

L368–369: "the simulation was slightly overestimated" -> "the simulation slighted overestimated"

L374: What about the LWP peak of ~100 g/m2 near C02 in HAMP? This is an order of magnitude larger than simulated LWP around that location.

L390: The fact that the P5 observations do not align with the air mass trajectory makes it questionable to compare them to the simulation.

Section 4.4: Is it possible to estimate the surface heat fluxes using the bulk method and compare them to the thermodynamic method? How do we interpret the estimated fluxes at steady state whereas the simulations are evolving drastically?

Figure 8 caption: the triangle legend is different from what's on the figure, which may cause confusion.

Figure 12: This figure took some time for me to grasp. What explains the deviation from the 1:1 line? I would appreciate seeing some profiles of buoyancy flux at different stages to better understand the timing of the decoupling.