General comments:

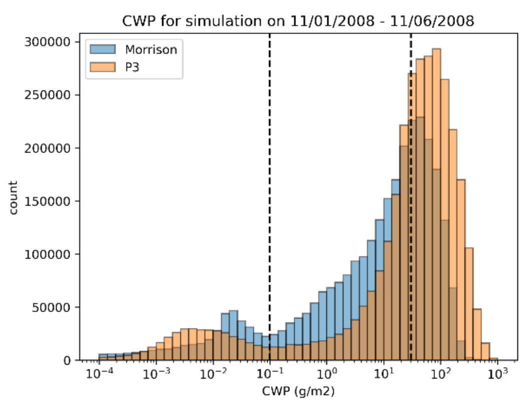
The study is well presented and nicely constructed. The findings are interesting. I suggest some minor/specific comments.

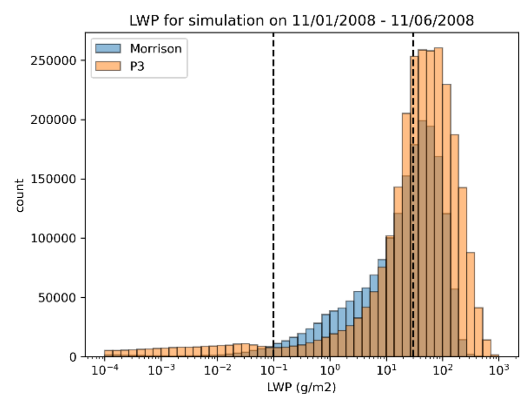
Specific comments:

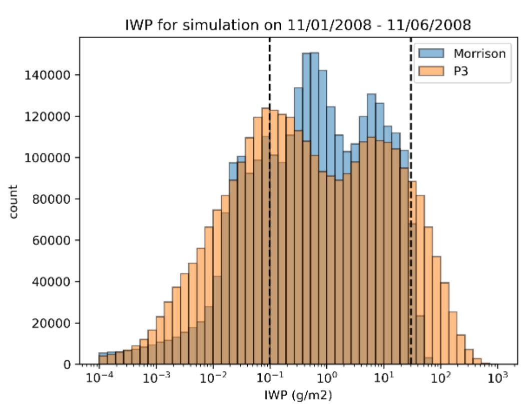
- L-63 : not in DJF, why?
    - ASR tends to struggle with producing a sufficient number of opaque clouds in general (see Bertossa et al., 2025). DJF is the season when the opaque mode is least prevalent in observations, as ice-covered surfaces and relatively low atmospheric moisture inhibit the formation of clouds with high water path. As a result, any underrepresentation of the opaque mode in ASR leads to an effective loss of the bimodal structure during DJF.
- Figure 2: How is the CRE obtained from the ground station? Is it the same definition as the satellite? (F_dwn – F_up)_all - (F_dwn – F_up)_clr ? The maximum of occurrence ARM-NSA Vs CSC does not occur at the same value of CRE-Lw (75W/m2 Vs ~65W/m2). It looks like the satellite retrieval occurs at higher values for all seasons and agrees the most in SON. Can the author explain why?
    - While it's the same basic formula, there are important differences with how each component is derived. The ground-based CRE relies on direct measurements of downwelling radiative fluxes (F_dwn) and the clear-sky (and CRE) fluxes are calculated from measured water path values. That is, F_dwn is known and F_dwn_clr is inferred. Whereas the satellite-based F_dwn (and CRE) is a derived quantity computed using a radiative transfer model that combines clear-sky flux estimates with retrieved hydrometeor properties. The close agreement between these two independent approaches provides strong confidence in the existence of the identified cloud states in the first place. Differences in the magnitude of the dominant CRE mode (i.e., 65 versus 75 W/m2) likely arise from a combination of factors, including the differing calculation methodologies, but also colocation differences. In particular, CSC values are based on satellite overpasses within a 1x1 degree box surrounding the station. Because the station is located near the coast, whether an overpass samples sea ice or open ocean can substantially influence the retrieved CRE, contributing to the observed offsets.
- L-175. Why is P3 overproducing opaque clouds while the Morrison one the intermediate mode? I mean, what is the microphysical configuration that causes

these discrepancies? Figure 6 shows that almost all the DLR clr have the same distribution across the different datasets/configurations and the differences are in the CRE-LW.

- This mainly comes from differences in cloud water path distributions. Below we've provided an example from a randomly selected simulation set. We've added vertical lines at 0.1 g/m2 (values lower than this are almost always transmissive according to observations, see Bertossa et al. 2025) and 30 g/m2 (values higher than this are almost always opaque according to observations). Values in between these lines can lead to "intermediate" CREs, though these don't actually occur frequently in nature. We see that Morrison has a greater relative frequency in this intermediate zone compared to P3. If we were to plot LWP and IWP specifically (below), we see it can be attributed to both phases. Beyond this, what physical mechanism causes P3 to produce more opaque clouds than Morrison hasn't been determined here. Our main goal for this study was to determine the scheme that best matches observations, rather than investigate the root mechanisms that lead to this. That being said, see our response to reviewer 2 #4 for a possible explanation to why P3 produces more opaque clouds.







- L-190: interesting because climate models usually underestimate opaque clouds. Wonder what in P3 scheme creates the overproduction of opaque clouds.
    - See previous response for a possibility.
- L-193. CALIPSO can get fully attenuated in opaque clouds, and CloudSat suffers from the strong echo from the surface that can alter low cloud detection over rough surfaces, but shouldn't be an issue over the Arctic. Is the issue the coarse vertical resolution of the satellites?
    - While this is a good consideration, the coarse resolution has little effect on the actual CRE in this case, so long as the cloud is detected in the first place. Cloud detection rates by CloudSat-CALIPSO product are very high in these cases (Bertossa and L'Ecuyer 2024). Henderson et al. (2013) find that the radiative influence of extending the cloud base has a relatively small effect (1.5 W/m2 / 250m). So even if the base can't be accurately captured due to attenuation or surface clutter, the effect on CRE should be relatively small. We have added note of this in the revised manuscript. This has been added to the revised manuscript (Lines 192-197).
- Figure 7: Did the author do some sensitivity studies on the vertical resolution of the radiative computation? Maybe it can explain some of the discrepancies. Or authors can add uncertainty bars for the different datasets and check whether they overlap.
    - Sensitivity to vertical resolution for the radiative computation is partly looped into Fig. 3 (p3 versus p3_hv). Where, increasing the number of vertical levels by ~50% does very little to change the resulting CRE distribution (both in terms of what clouds are produced, and how the radiative effect of those clouds are calculated). We have added explicit mentions of the lowest model level for each vertical configuration to the revised manuscript (Fig 3 caption and model configuration table).
- Figure 8: The mean cloud fraction profile in the Morrison scheme (over ice) shows a bimodal distribution with a large population of low clouds (~500m) and mid-level clouds (5-6 km). I don't understand why we don't observe a bimodality in the CRE-LW distribution that follows the cloud fraction one. Could the authors clarify this point?
    - This is great to point out, and part of the reason why doing these model evaluation studies can be difficult. "Clouds" in models and clouds detected

by observations are not necessarily equivalent. “Clouds” in models are often defined by a non-zero water path value, which doesn’t necessarily have a significant radiative effect. Furthermore, observations do indicate that there are preferred water path values (see Bertossa et al. 2025 Fig. 9&10). That is, there is a physical instability that pushes the state towards near-zero water path values or very high water path values that cause the CRE to saturate. There are a lot of great suggestions as to why this may be the case (see for example Morrison et al 2012 and references therein). This plot indicates that the Morrison scheme does not appropriately lead to these preferred water path groupings as we see in observations (and P3). See discussion in Lines 240-250.

Figure comments:

- Using green and red in the same plot is not color blind friendly. I suggested changing the line style when these two colors are used.
    - Revised as suggested. Green bars have been replaced with black.
- In fig. 6, the legend says ‘Obs’ for the black line. I suggest kipping CSC for satellite retrieval and ARM-NSA for the ground station (or similar) throughout the paper. Same for the simulations, keeping the same notation can be helpful.
    - Revised as suggested. Obs has been replaced with CSC here.