Comments:

Line 1: The title specifies this study as "Cloud properties over the Southern Ocean During the MARCUS field campaign". The "cloud properties" is a broad description including macrophysics, microphysics and dynamics, etc. However, the focus of this work is on the investigation of the macrophysics distribution and the cloud phase characteristics. I suggest using a more specific title to be consistent with the scope of the manuscript.

Line 17: It seems that cloud type is classified with ambiguity. Low-level clouds are identified based on the cloud altitude, while shallow cumulus clouds are classified with their morphology and height. I would prefer to consider the shallow cumulus are a subgroup of low-level clouds instead of a separated group. In the manuscript the authors only use cloud height to identify the shallow/deep cumulus, which is not sufficient. I suggest a thorough revision of these terminologies and use them consistent in the manuscript to avoid any confusion. Specifically, I suggest using low/middle/high clouds and avoid the "cumulus" category unless a more specific cumulus classification algorithm is being added in the manuscript.

Line 134: Does the cloud occurrence frequency shown in Figure 2 sensitive to the length of the time window? If so, why use 5 min? Line 188 also mentions the relative low cloud fractions compared to Satellite results is due to the inconsistent resolution. Adapting the wind speed from sounding, a reasonable "equivalent temporal resolution" of the surface-based radar may be obtained. For example, considering the windspeed as 10m/s, a 2-grid box(~200km) corresponding to approximately 5.5 hours of the radar observation (not considering the speed of ship). Thus, does the cloud frequency estimated from a longer time window is closer to the spaceborne results? I'm not suggesting performing a sensitive test for Fig.2, but just want to remind the authors that it is necessary to discuss this issue furthermore for the presented results to be better used for model validation, as the authors mentioned in the introduction.

Line 200: How close? What is the specific CF from Listowski et al, 2019?

Line 214: I also mentioned in the previous comment. The classified cloud type is solely based on cloud heigh, it is not clear why the MOL is related to shallow cumulus.

Line 215: Are you refereeing to occurrence frequency is higher?

Line 226: For Figure 3. Why use the deepest cloud thickness instead of the mean/median thickness? It is hard to interpret the thickness distribution in a statistical perspective as the deepest thickness may be the results of two "outliers" of the cloud in each group. Besides, I suggest using the box plot to present the results shown in Fig. 3b.

Line 232: How the stability is estimated and what is the value for NSO?

Line 235: …HGH and HOM peak "are" less than 10 g m^{-2}

Line 241: Same as before, the shallow convective clouds are not clearly identified.

Line 248: Table1: Compared to the mean LWP, the std (uncertainty) is too large and indicates the result is statistically insignificant. This may be due to the insufficient MWR observations being collected in the analysis. Have you considered the period when MWR is wet and thus the LWP retrieval becomes unreliable? The maximum LWP value (1937 g m^{-2}) seems too large to be accurately retrieved by MWR.

Line 250: Does the "non-contiguous clouds" refers to multi-layer clouds?

Line 251-252: Please add the corresponding statistics in Table 1 to support this statement.

Line 266: Same as before, I suggest using the mean cloud thickness instead of the "maximum" cloud thickness before making further statements.

Line 288:290: What is the meaning of the "majority of WID?" Why not using the mean/median WID? What is the radar temporal resolution?

Line 292: And or "or"?

Line 289L Any supporting information regarding on the chosen of 0.4m/s of WID?

Line 288-Line 294: Please rephrase the description of the classification algorithm, which is not consistent with Figure 5.

Line 293: LWP threshold (10 gm^{-2}) from the manuscript and Figure 5 (20 g m^{-2}) is not consistent.

Line 305-310: Not quite understand why this analysis is helpful to validate the proposed classification algorithm.

Line 315: I cannot tell the backscatter difference for Fig.6e, please adjust the color bar range to make the figure more distinguishable. Also, the backscatter unit is missing.

Line 321-323: From Fig. 6c, the identified ice period from radar around 22 UTC corresponding to low LDR value in Fig 6.d, which seems contradictory to this statement.

Line 337-339: It is necessary to present the SOCRATES results if this can further support the classification output. One thing should be considered for the onboard radar is that the Doppler spectrum width may be much higher than the surface-based radar as for the enhanced Doppler spectrum broadening caused by the moving of aircraft.

Line 345-346: What is the difference between mixed-phase and "other" clouds?

Line 364: Drizzle is one kind of the liquid water droplets; why drizzle occurs in ice cloud?

Line 411-Line 413: This comparison only make scene if the cloud phase classification algorithm is same for these two studies.

Line 417: Radar reflectivity need to be calibrated before used for analysis. (Mace., 2021)

Line 429: This statement is valid only if the vertical air velocity is negligible compared to the terminal velocity of cloud/ice particles.