Response to Reviewer 2:

Overall comments:

This manuscript evaluated cloud radiative biases in the ACCESS-AM2 model and the CERES satellite product against ground-based observations during the MICRE campaign at Macquarie Island. The study used a suite of instrumentation to establish a positive bias in surface shortwave radiation in both the model and CERES, which is consistent with numerous past radiative flux modeling studies over the Southern Ocean. The authors also employed a new lidar simulator to constrain model output with observational instrument specifications, therefore allowing an evaluation of cloud attribution to the radiative biases. They found that excessive absorbed surface shortwave radiation is associated with too low of a cloud fraction. Overall, the manuscript is well-written, descriptive where needed and brief where appropriate, and the experiment is well-designed. The largest concern is the lack of discussion of the potential role that cloud microphysics have in the discussed biases. While I understand this is not a microphysics study, there are long-standing microphysical biases in GCMs that are suggested to be relevant to simulating Southern Ocean clouds (e.g., prevalence and maintenance of supercooled liquid) that could be discussed in more detail as potential caveats. Otherwise, I have provided some general comments and line-specific comments below, none of which should require too much extra effort to delay publication. I suggest this manuscript be accepted for publication after these minor revisions are addressed.

We thank Reviewer 2 for taking the time to comment on our manuscript. We have provided more discussions about the potential role of cloud microphysics have in the radiation biases. More responses for comments are as below.

General comments:

1) ***In the second paragraph of the Introduction when you introduce the ALCF, I think some expansion is needed for unfamiliar readers of how this framework operates. Typically, simulators are thought of as being applied to model data, not on observational data, so this can be confusing and deserves a little more explanation. Based on your description in Section 2.6, it seems ALCF is used on the observational ceilometer dataset as more of a means of calibration rather than "simulation".***

Indeed, ALCF is not used as a simulator with observational data. It performs calibration, resampling, and cloud detection, in a way which is corresponding the processing of model data. We have added more description of ALCF in the second paragraph of the Introduction and Section 2.6.

Line 43: This is accomplished by extracting two-dimensional profiles (time x height) from the model data, using a modified COSP lidar simulator to perform radiative transfer calculations, calibrating and resampling the observed attenuated volume backscattering coefficient to a common resolution, and conducting similar cloud detection on both the simulated and observed attenuated volume backscattering coefficient (Kuma et al., 2021).

Line 240: For the model data, ALCF first extracts two-dimensional cloud liquid and ice content profiles at the survey area, then uses Subgrid Cloud Overlap Profile Sampler (SCOPS) to generate 10 random subcolumns for each profile to detect clouds in the model (Chepfer et al., 2008). The default setting for generating cloud overlap is maximum-random overlap assumption, which assumes neighboring layers with non-zero CF are fully overlapped, while layers separated by zero CF are randomly overlapped. The same sampling rate (5 min) and vertical bins (50 m) were used in lidar simulator to make the model and observations comparable. The attenuated volume backscattering coefficient profiles are then simulated for 10 subcolumns based on the COSP lidar simulator. Subsequently, ALCF re-samples the observational profiles to increase the signal-to-noise ratio, subtracts the noise, calculates the lidar ratio, applies an absolute calibration, and uses a cloud detection algorithm to calculate cloud mask and CBH for both simulated and observational data. A threshold of 2×10^{-6} m^{-1} sr^{-1} for backscattering coefficient is applied to identify cloud mask, as this value was found to be a good compromise between false detection and misses in Southern Hemisphere, where the data is less impacted by anthropogenic aerosol. This step is important to make the simulated and observed backscattering coefficient profiles comparable.

Line 258: Several limitations exist within the ALCF that can cause uncertainties (Kuma et al., 2021). Firstly, the accuracy of the CL31 and CL51 ceilometers' calibration may be impacted by the absorption of water vapour at 910 nm, which can limit the precision of their comparison. However, it is improbable that the calculated cloud masks will be significantly influenced due to the high backscattering caused by clouds. Secondly, precipitation and aerosol are not currently implemented in the simulator. The cloud detection algorithm typically identifies observed precipitation as "cloud", whereas the simulated profile does not show any backscattering in the area where precipitation is occurring. Upon reviewing the backscatter profiles, certain layers beneath stratocumulus clouds are identified as clouds, potentially consisting of drizzle, snow, fog, or aerosol. Nevertheless, the frequency of such occurrences is insufficient to significantly impact the statistics in a manner comparable to the model bias. Finally, the ALCs also encounter several measurement limitations. Specifically, inadequate overlap, dead time, and after-pulse corrections often yield sub-optimal outcomes in the close range. Semi-automated methods include calculating the distribution of integrated attenuated volume backscattering coefficient by analyzing the height where maximum backscattering occurs.

2) ***In the fourth paragraph of the introduction, I think something should be said about satellite limitations in observing low-level clouds over the SO, which is a very strong motivation for evaluating ground-based observations. As one example, Tansey et al. (2022; https://doi.org/10.1029/2021JD035370) looked at surface precipitation measurements during MICRE in comparison with CloudSat, with a few notable differences based on satellite instrument sensitivities and algorithm structure.***

   We have added statements for satellite limitations as follows.

   Line 73: 'Tansey at al. (2022) examined surface precipitation measurements during MICRE and compared them with data from CloudSat, revealing several notable differences attributable to satellite instrument sensitivities and algorithm structure. This indicates the limitations of satellite in observing low-level clouds over the SO, which serves as a strong motivation for utilizing ground-based observations to calibrate satellite products.'

3) ***You discuss how observed cloud fraction is defined by the all-sky cloud camera (Section 2.2.3) and then state it is used to evaluate CF biases on line 337 in Section 4, but it's not entirely clear how CF is being defined by ACCESS-AM2. Is it the prognostic value? Is it the CF computed by ALCF? It would be good to mention this at the beginning of Section 4.***

   We have provided how CF is defined by ACCESS-AM2 in Section 2.4. It is the prognostic value from PC2 cloud fraction scheme.

   Line 221: 'Of interest to this study, the ACCESS-AM2 model uses the Suite of Community RAdiative Transfer codes based on Edwards and Slingo (SOCRATES) (Edward and Slingo, 1996) and Wilson et al. (2008)'s prognostic CF and condensate cloud scheme, which includes large-scale as well as convective clouds. For comparison with the observational data, radiation and prognostic CF in the model was linearly interpolated to the point nearest to Macquarie Island (54.5°S, 158.9°E).'

4) ***Section 2.1 and other parts of manuscript–I wouldn't call this in-situ observations, since many of what is included are more commonly thought of as passive remote sensing instruments. Perhaps change to "ground-based" observations, or convince me otherwise.***

   We have changed all 'in-situ' to 'ground-based'.

5) ***Figure 1 caption–what does the blue color scale represent in Fig. 1a? I'd briefly mention it in the caption.***

We have mentioned in the Figure 1 caption: 'The blue color scale represents the bathymetry of oceans.'

6) ***Data availability: There was no data source given for the University of Canterbury ceilometer (CL51) data. Please include it as follows: https://doi.org/10.26179/5d91835e2ccc3 . In addition, no data sources were given for the radiometers or the all-sky cloud camera. The ARM data availability statement suffices for the ARM instruments, but if a DOI exists for these AAD instruments, they should be listed too.***

We have added the data source for UC ceilometer in the *Data availability*. Data sources for AAD radiometers and cloud camera will soon be published on Australian Antarctic Data Centre (AADC).

Line 609: 'The AAD radiometer and all-sky cloud camera data will shortly be available from the Australian Antarctic Data Centre (AADC). The University of Canterbury's Vaisala CL51 ceilometer data are available at AADC (https://doi.org/10.26179/5d91835e2ccc3).'

7) ***Perhaps a *little* more could be said about cloud phase in the last 2 paragraphs of Section 6 and how more or less supercooled liquid in the model relative to observations can impact your results. In general, the discussion of cloud microphysics in the manuscript is rather weak, and providing some speculative pathways for explanation would be very helpful.***

We have emphasized SLW's role in impacting the radiation biases in Section 6.

Line 600: 'We emphasize that the correct representation of supercooled liquid water over the SO is important for modelling the radiation in the region, as inadequate supercooled liquid water content will cause less reflectivity of clouds and result in positive downwelling surface SW biases (Luo et al., 2016; Vergara-Temprado et al., 2018; Gettelma et al., 2020).'

Moreover, we have provided more discussion of cloud microphysics in the last paragraph of Section 5, which starts from Line 531.

Line-specific Comments:

1) ***Line 108: "was" should be "were"***

Word has been changed.

2) ***Line 110: The two clauses should be joined by a conjunction, not a comma.***

   Sentence has been modified.

3) ***Line 136: I think you have to be careful when saying that supercooled clouds are typically not visible in the backscatter profile. Liquid-based supercooled clouds can exist at rather low altitudes (< 1-2 km AGL) over Macquarie Island, and in the absence of an underlying layer, will show a sharp gradient in attenuated backscatter consistent with a liquid cloud base identification. I would also mention that fog can frequently be observed in the backscatter profiles.***

   We have deleted the 'supercooled cloud layers' in this sentence and added that fog can also be observed.

   Line 151: 'Information on CBH, precipitation, and infrequently boundary layer height can be obtained from the backscatter profile using detection algorithms. Fog can be observed in the backscatter profiles as well.'

4) ***Line 216: Please give some reference to what type of CBH algorithm is used, since these are not trivial nor converged methods. I assume it is what is described in Section 5.3 of Kuma et al. (2021), and if so, I'd list the thresholds for attenuated volume backscattering coefficient that were employed here.***

   We have provided the information of thresholds for attenuated volume backscattering coefficient in Section 2.6. The threshold we used is 2×10^{-6} m^{-1} sr^{-1}, which was found to be suitable for the Southern Hemisphere.

   Line 248: 'A threshold of 2×10^{-6} m^{-1} sr^{-1} for backscattering coefficient is applied to identify cloud mask after removing 5 standard deviations of range-scaled noise, as this value was found to be a good compromise between false detection and misses in Southern Hemisphere, where the data is less impacted by anthropogenic aerosol.'

5) ***Lines 270-271: This statement is very confusing to me. I'm not sure what you mean by large spreads in the LW_cs distributions, as the violin plot in Fig. 3d doesn't really show spread that is larger in model/satellite relative to ERA5. Also not sure what you mean by "paying more attention to the LW_cs models than the SW_cs models". Are you suggesting that LW_cs biases are more important than SW_cs biases, or appear to have a higher sensitivity? Please clear this up.***

   Here we tried to emphasize the larger biases in LWcs and that more work is needed in this space. We have modified the sentences as follows.

Line 318: ‘The significant differences of LWcs in model and satellite compared to ERA5 highlight the need for more validation on development of especially the LWcs models. The SWcs models show smaller and insignificant biases, indicating less uncertainty.’

6) ***Line 275: Careful with this statement. “Too few, too bright” refers to compensating errors. All else being equal, “too bright” should reflect more surface SW radiation, so this component can’t really be linked to an overestimation of absorbed surface SW.***

Yes, too bright clouds will result in underestimated surface SW radiation. We have revised this statement as follows.

Line 327: ‘"Too few and too bright" low-level clouds were identified as the cause of this SW bias in CMIP5 models (Nam et al., 2012; Wall et al., 2017). Nevertheless, more recently, Schuddeboom and McDonald (2021) discovered the exact contrasting result in the CMIP6 simulations, which demonstrates the importance of prioritizing the low-level cloud simulation to enhance the SW radiative balance over the SO.’

7) ***Lines 289-291: Aside from the absolute values being ~ 2 W/m^2 off, the values you listed do seem consistent with Hinkelman and Marchand (2020)---SW bias of +8 for your study and +10 for their study, and LW bias of -12 for your study and -10 for their study. Understanding that 2 W/m^2 is not a small amount, I’d suggest rewording because “consistency” usually refers to how biases operate in sign, even if magnitudes are different.***

We reworded the ‘not consistent’ to ‘not equal to’. Moreover, we have added more speculation for this difference between two study.

Line 344: These differences in SW & LW biases are possibly attributed to different temporal resolution of the CERES SYN product (hourly output used in Hinkelman and Marchand (2020) and daily output used in this study) and different interpolation methods to collocate data to Macquarie Island (Hinkelman and Marchand (2020) chose the nearest grid that contains Macquarie Island while this study linearly interpolated data to Macquarie Island). Other factors such as data gaps, sampling uncertainty, calibration offsets, different pyranometers, and local shadowing effects may also contribute to the biases difference.

8) ***Line 299: It’s not necessary for this study, but perhaps the frequent soundings released at Macquarie Island can give you an idea of potential humidity and temperature biases in ERA5.***

The soundings data at Macquarie Island has been assimilated into the ERA5 product, thus we expect that the ERA5 data performs satisfactorily compared to the sounding observations at Macquarie Island.

9) ***Line 302: It's not clear to me what you mean by three algorithms used for SW_cs biases. Are you including the ERA5 SW_cs calculations in this statement?***

   Here, we meant ACCESS-AM2 and CERES used similar algorithms to ERA5. We have corrected the statement.

   Line 359: 'Here we have shown that while the SWcs biases from ACCESS-AM2 and CERES (using similar meteorology driven by ERA5 and using the same method of calculating the clear-sky fluxes) are very similar, the same cannot be said for the LWcs.'

10) ***Line 304: Again, I wouldn't use "in-situ" here, as I think much of the community thinks of these as passive remote sensing instruments.***

    We have changed 'in-situ' to 'ground-based'.

11) ***Lines 301-305: This paragraph seems a little out of place as you start discussing CREs. I expected this to be a transition to the next section, but that also wasn't clear because the paragraph ends on a "future guidance" type of statement. Suggest making the transition more clear or moving this paragraph to the next section.***

    We added more discussions of ERA5 radiation biases.

    Line 356: Understanding the biases in the respective SW and LW clear-sky biases is an important but often neglected component of understanding the CREs. Here we have shown that while the SWcs biases from ACCESS-AM2 and CERES (using similar meteorology driven by ERA5 and using the same method of calculating the clear-sky fluxes) are very similar, the same cannot be said for the LWcs. These differences, and how they affect the CRE, require further study. Wang et al. (2020) evaluated the cloud radiative effect of ERA5 using ship-based measurements in the SO during three summer seasons. Higher shortwave cloud radiative effect (+77 W m^{-2}) and lower longwave cloud radiative effect (-18 W m^{-2}) were detected in ERA5 in all-sky conditions, which are likely attributed to the higher occurrence of clouds over the Southern Ocean compared to what was modelled, and potentially resulting from the higher transmittance of clouds in the ERA5 (Wang et al., 2020). Regarding clear-sky conditions, no notable error was found in the ERA5 LW irradiance, while for SW, the observed values were 33 W m^{-2} higher than those predicted by ERA5. More recently, Mallet et al. (2023) found large downwelling SW radiation biases (+54 W m^{-2}) in the ERA5 compared with 25 years summertime surface measurements collected from ship and ground station over the SO. By employing machine learning techniques, cloud cover and relative humidity exhibited a strong contribution to the SW radiation biases. Despite these few studies on ERA5 radiation biases, a limited amount of research has been dedicated to investigating this issue, particularly in relation to clear-sky conditions. We suggest the importance of using ground-based observations of clear-sky radiation to evaluate the model and satellite, as well as validating the reanalysis product.

And we added a paragraph for transition as follows.

Line 374: 'After investigating the SW and LW radiation biases of ACCESS-AM2 and CERES in both all-sky and clear-sky conditions, we next assess their capability to reproduce cloud radiative effect.'

12) ***Line 316: You haven't yet used the term "downward CRE" and it's not used anywhere else in the manuscript, so I'd avoid it here to avoid confusion.***

We have removed 'downward'.

13) ***Line 323: You say larger negative values of CRE_LW, but CRE_LW values are strictly positive using your convention.***

Here, we meant the larger negative biases. We have changed 'values' to 'biases'.

14) ***Line 355: Fig. 6e does not show an overestimation in CF in spring as stated here, but rather the same mean CF for ACCESS-AM2 and observations.***

Yes, we have deleted 'spring'.

15) ***Line 360: Is there a reason for suggesting cloud microphysics are a lesser control on the CRE compared to other properties? I would add some references to back this up if so. Droplet radius and size distributions are inherently linked to the cloud's optical thickness, after all.***

Cloud microphysics also plays an important role in affecting radiation biases. We have modified the sentence and provided more references as follows.

Line 430: 'Nevertheless, the overall overestimated CF and positive surface SW biases in the model indicate that the CF alone does not control the cloud radiative effect, but also properties such as cloud phase, cloud base height, and cloud geometrical or optical thickness are likely to play a significant role (Viudez-Mora et al., 2015; Cesana and Storelvmo, 2017; Fiddes et al., 2022). In addition, cloud microphysics such as ice crystal habit and size distribution and direct and indirect effect of aerosols could also have an effect on radiation biases (Bohren and Huffman, 2008; Kuma et al., 2020). Our results here are in agreement with the work done by Schuddeboom and McDonald (2021), which found overestimated low-level CF and reduced reflectivity of low-level cloud over the SO in CMIP6 models, highlighting the significance of correctly representing low-level clouds to simulating radiative balance over the SO.’

16) ***Line 382: ”results” should be singular “result”***

Word has been changed.

17) ***Line 385: Suggest using “While much of that is...” instead of “While a lot of that is...”***

Sentence has been changed as suggested.

18) ***Line 473: Saying “other” CFs is rather vague. I would provide a little more detail here and give the range where the model produced lower CF (between 0.2 and 0.6), or just say “lower” CFs instead.***

We have specified more details about ‘other’ CFs.

Line 567: ‘However, this is caused by an underestimated frequency of CF between 0.2 and 0.6 and an overestimated frequency of CF above 0.6.’

19) ***Line 484: I think you mean under positive SW *bias* conditions. Also, “result” should be “results”.***

Words have been changed as suggested.

20) ***Line 488: Suggest making it clear that you are referring to a greater dependence on low-level CFO for LW biases *compared to* SW biases, if that is indeed what you mean.***

Yes, we have modified the sentence as follows.

Line 582: ‘We suggest that the larger differences in the modelled low-level cloud occurrences between the LW conditions demonstrates the greater dependence on low-level CFO of the LW biases than SW biases.’

21) ***Line 489: I wouldn’t say “inappropriate” cloud representation. It’s as appropriate as can be given scale separations, but must inevitably be parameterized. I’d suggest using “parameterized” instead of “inappropriate”.***

We reworded ‘inappropriate’ to ‘erroneous’ as suggested by Reviewer 1, as we want to stress the incorrect cloud parameterization in the model.

22) ***Line 499: Again, wouldn’t use “in-situ”***

We have changed ‘in-situ’ to ‘ground-based’.

23) Line 523: Should be either “These analyses suggest” or “This analysis suggests”.

Sentence has been changed as suggested.