

Pei et al. compare surface observations of downwelling shortwave and longwave radiation at Macquarie Island to the ACCESS-AM2 model and CERES product. They then analyze radiative biases as a function of cloud fraction and cloud occurrence, using the ALCF simulator (relying on hourly output from the model) and an all-sky cloud camera for the observational component. Overall, the authors clearly relate respective CRE biases to the identified shortcomings of both ACCESS-AM2 & CERES, in comparison to the MICRE observations. They establish a strong relationship between cloud fraction and SW/LW at the surface in both the model and observations. Interpretation of the role played by cloud occurrence (especially low clouds) is weaker, and I recommend that the authors provide more background information on the role of cloud microphysics (beyond cloud phase). This may include a brief survey of studies of Southern Ocean cloud properties from surface campaigns and in-situ aircraft measurements (e.g. during SOCRATES). I also recommend that the authors report the fraction of (and reasons for) missing/bad data for *all* observations (and whether you expect these to introduce any sampling bias), and more detail on the limitations of using ERA5 for clear-sky radiation estimates. I recommend that this paper be published after minor changes.

General comments

- 1) In the introduction section second paragraph, more could be said about the sparseness of surface-based SO ceilometer observations (cloud occurrence frequency and cloud boundaries) and sources of uncertainty within the ALCF data. You state that this is a recently developed tool. Include a brief description of the ALCF here (with more detail in methods).
- 2) Regarding uncertainty, data issues and missing data fractions for all surface-based measurements: You provide a good description of this for the radiometer. For the MICRE ceilometer, can you provide a % of time that you do not have good data (missing a successful cloud base retrieval) because of fog, snow on the detector or other reasons for missing data? How does fog and attenuation affect your CFO profiles? (See also "general comment" #4 below.)
In reference to the all-sky cloud camera, can you report the uncertainty and bad/missing data %?
- 3) In methods section 2.3.2, you state that you're able to contain clear-sky radiative biases in equation 4 by relying on ERA5, which ACCESS-AM2 & CERES also rely on. But this re-emerges later in the paper as a driver of uncertainty in the interpretation of results. Are you aware of any literature worth mentioning (whether early in the paper or in the discussion section) that provides more information on ERA5 radiative biases over the region? I scanned briefly and found one very recent article, for example:

Mallet, M. D., S. P. Alexander, A. Protat, and S. L. Fiddes, 2023: Reducing Southern Ocean shortwave radiation errors in the ERA5 reanalysis with machine learning and 25 years of surface observations. *Artif. Intell. Earth Syst.*, <https://doi.org/10.1175/AIES-D-22-0044.1>

- 4) Figures 8 & 9 are of concern because of the very large CFO near the surface. Describe in detail how you got this profile and how you calculate cloud fractions at each height. Especially near the surface, >50% cloud occurrence in the lowest 50 meters seems improbable. If your approach was to define a minimum threshold on the ceilometer backscatter, you should discuss (and potentially rethink) this.
- 5) In the interpretation of results in section 5, I think you should discuss other possible sources of radiative biases here (e.g. differing cloud macro and microphysical properties). It is not sufficient to state only that the low-level CFO was less negative for negative SW biases, and provide no additional reasoning. If two opposing extremes of SW biases are both happening with underestimated low-level CFOs, additional cloud properties must play a role here.
Lines 502-505 start to address this. I think the discussion (and the paper overall) would be strengthened by referencing additional studies of cloud radiative effects as a function of macro and microphysics.
- 6) You have two years' worth of seasonal cycles, and yet you do not discuss the differences in the two years. It would be interesting to see how much downwelling radiation, cloud fraction and vertical occurrence frequencies varied in each season during 2016-2017 as compared to 2017-2018.
- 7) Appendix B: What about diurnal errors in the CERES SYN product?

Line-specific comments

- 1) Line 50-52: You should also acknowledge here that discrepancies even exist between surface and satellite observations, due to limitations of satellite near-surface cloud retrievals.
- 2) Line 55-56: I suggest connecting cloud feedbacks to the previous sentences; what cloud feedbacks are relevant to this study?

- 3) Line 62-63: Suggest rephrasing “remote atmospheric environment” to “the harsh atmospheric environment and lack of remote sites for measurements”
- 4) Line 99-101: This explanation makes it difficult to follow what the time resolution is. 1-minute means and standard deviations? I suggest stating this up front.
- 5) Line 104: I suggest writing out “microvolts”
- 6) Line 107: Please describe the calibration process in more detail. How did you get these coefficients? How do you determine sensitivities?
- 7) Line 119: Consider relating the 0.5°C to a % error in the measurements. Presumably this will have a very minimal effect on equation (1) since T_b is in Kelvin, although it is raised to the fourth power.
- 8) Line 123: “*limited* clipped points”
- 9) Line 127-129: In my opinion this should be stated earlier in the section.
- 10) Line 133: Do you decrease 6-second sampling to 1-minute resolution to match up with other data? If so, what is your approach (such as using the median CBH, do you require a minimum detection threshold i.e. X columns out of 10 columns in the minute contain cloud base retrievals?)
- 11) Line 190: Explain why September-February were chosen to look at hourly simulations.
- 12) Line 212-214: Provide some comment on the input to the ALCF simulator.
- 13) Line 217: misspelled ceilometer
- 14) Line 218-219: What is the spacing of height bins? What is a column? E.g. a 1-minute vertical profile of 100-meter vertical bins extending up to 15km?
- 15) Line 233: “Generally consistent” - I suggest making this statement quantitative: ACCESS, CERES match observations to within X,Y in W/m^2 , “as seen in table 1”. You start to do this at line 250, but a comment on general consistency can be made earlier to prepare the reader for what constitutes a significant disagreement.
- 16) Line 237: Curiously, this underestimation seems more pronounced in the second winter (JJA 2017) when I look at Figure 2. Is this the case? If so, perhaps comment on this in the text?
- 17) Line 254-255 (and re: Figure 3 in general): Perhaps add a sentence here or in the figure caption explaining the violin plot spread at various radiative fluxes (vertical axis values), and what the bolded segment is meant to show on the middle line.
- 18) Line 256: I suggest a rewrite, “reaching $-4 W/m^2$ in Autumn, with smaller differences in all other seasons.”

- 19) Line 275-276: In more recent generations this “too few, too bright” issue has been modified -- perhaps talk about findings from Schuddeboom and McDonald (2021). Are the cloud fraction differences discussed in agreement with this analysis?
Schuddeboom, A. J., & McDonald, A. J. (2021). The Southern Ocean radiative bias, cloud compensating errors, and equilibrium climate sensitivity in CMIP6 models. *Journal of Geophysical Research: Atmospheres*, 126, e2021JD035310.
<https://doi.org/10.1029/2021JD035310>
- 20) Line 289 You write that the biases found here (SW $+8.0 \pm 18.0\text{Wm}^{-2}$, LW bias of $-12.1 \pm 12.2\text{Wm}^{-2}$) are “NOT consistent” with the previous study by Hinkelman and Marchand (2020) (SW $+10\text{Wm}^{-2}$, LW -10Wm^{-2}). How so? The means would appear to be nearly the same, and well within the listed uncertainty?
In general, I don’t understand the \pm values given here, as they don’t seem to match table #1.
- 21) Line 292: Isn't the ARM pyranometer from Hinkelman & Marchand [2020] the one used for the comparisons in Figs. A1 & A2? One would think calibration offsets and local shadowing effects have only minor impacts since the comparisons in the appendix are in such good agreement.
Also earlier you said only 9 days have been removed from your timeseries. So are data gaps in the ARM pyranometer more frequent?
- 22) Line 296. You write “Excellent alignment of SWcs radiation between the satellite and reanalysis is expected given the CERES product uses ERA5 to inform its radiative transfer algorithm.” Perhaps start a new paragraph as this represents a change in topic.
- 23) Line 312-313: I would argue from downwelling SW in Fig. 3d that the winter and spring clear-sky LW are somewhat lower, exhibiting a weak but discernible seasonal cycle. Suggest removing "clear-sky" or changing to "similar to the all-sky and, to a lesser degree, clear sky"
- 24) Line 315: “Figure 5 *shows*”
- 25) Line 319: “During winter, when total CRE is at the lowest value” – JJA is the season where the CRE is highest (the most positive value). And if you are talking about magnitudes only, it is lower in MAM than in JJA. I’m not sure what you’re referring to here.
- 26) Line 325: Suggested that you change “climate models” to “ACCESS-AM2”
- 27) Line 329: “which *peaks*”
- 28) Caption of Figure 6 & line 341: Use the plural, “*observations*”

- 29) Line 342: Draw the reader's attention to specifically Fig. 6a to demonstrate the integrated effect discussed in this sentence.
- 30) Line 343: Only Figure 6e is for spring.
- 31) Line 350: "mid-level CF": You might include the thresholds here (above 4 km / below 10 km or something).
- 32) Line 357: "*particularly*"
- 33) Line 360: If you state that microphysics plays a lesser role in cloud radiative effect than the macrophysics described in the previous sentence, you should back this up with references. More references in general in lines 358-363 would be good.
- 34) Line 369: "*thresholds*"
- 35) Line 374: Period after "LW bias" should be a comma
- 36) Line 382: "This *result*"
- 37) Line 388-389: Over what number of hours are you calculating cloud fraction? How are you averaging and comparing these two different datasets with different resolutions?
- 38) Line 389: "*Figures*"
- 39) Figure 8: I recommend adding shading around the vertical cloud fraction profiles indicating e.g. standard error on the mean CF at each height bin.
- 40) Line 395: "overleap" – you probably mean overlap
- 41) Line 404: "summertime only" – The period is spring and summertime.
- 42) Line 404: I suggest rephrasing to "which the model underestimated by 6%."
- 43) Line 405-406: Look for references on how often multi-layer clouds occur. This can give you some estimate of how frequently the ceilometer suffers this drawback of high cloud obscurity.
- 44) Line 408: I'm not sure I follow this. Why would ACCESS-AM2 output passed through ALCF suffer the same backscatter attenuation limitation?
- 45) Line 422-425: Perhaps include a range or estimate of the uncertainty on the 4% higher total CF. In any case, especially considering the missing low-cloud occurrence < 1km, this result is surprising.
Additionally, I cannot tell by looking at 9a that the red line yields 4% higher CFO than the black line, all vertical levels considered... But this could just be an effect of the vertical axis log scale.
- 46) Line 425: I suggest changing "Figure 8" to "as seen in Figure 8 and replotted in Figure 9 in lighter colors"
- 47) Line 428: "*model's*"

- 48) Line 433: I suggest changing “simulates” to “overestimates”
- 49) Line 439: Awkward and needs rewording. I suggest changing to "Below 500m, the modeled cloud occurrence is lower than the average,"
- 50) Line 444: “can be in the model”? Maybe reword this.
- 51) Line 446: “cloud phase” – Cloud microphysics in general are likely contributing to SW radiative biases, including cloud droplet size and number concentration, cloud optical depth.
- 52) Line 470-471: Given the dependence on ERA5, can you speculate on the source of this problem?
- 53) Line 472: “*observations*”
- 54) Line 472-473: Elaborate more on these compensating errors as a separate sentence.
- 55) Line 473: Remove the comma after “prior”
- 56) Line 489: “inappropriate” is a funny word choice; I suggest inadequate or erroneous.
- 57) Line 490-491: If data are missing during these periods, what makes you say there were outliers in the observations?
- 58) Line 491-492: “which will include inaccuracies” – Such as?
- 59) Line 508: “*by further*”