Response to Reviewer 1

Many thanks to this Reviewer for their careful comments. We appreciate the time they have spent highlighting where our methods and ideas were not clearly communicated. In particular, we thank the Reviewer for the suggestion of a diagram to help communicate our work. We have now resolved some of the questions this Reviewer had around our methods. Our detailed response can be found below.

Overall comments

The study presents a k-means cluster analysis of the recent ACCESS model in the spirit of (Williams and Webb 2009) and (Haynes et al. 2011). The paper could do more to concretely link biases in cloud properties to biases in radiation- radiative biases are suggested to be related to biases in various cloud properties, but this appears to be by eye rather than quantitative. In several places the writing is difficult to follow and especially in the analysis section it is hard to follow whether cloud RFO or cloud properties are being referred to (in several cases clouds are referred to as being simulated correctly or incorrectly, but it is unclear what that means in the regime framework) and often the ability of the model to replicate these quantities is described in vague relative terms. It is also not clear if the authors are comparing in-cloud and area-averaged water paths.

We have carefully read through the manuscript to make sure we are consistent with our terminologies (correct/incorrect definitions) and that we remove any potentially vague statements. Examples of this are specifically outlined in the remainder of the comments below.

Abstract: In several places the authors refer to incorrectly or correctly simulating clouds. It is ambiguous what they mean by this. It seems to be only referring to phase and frequency (which I think it equivalent to cloud fraction). If this is the case, it would be good to clarify that we only care about phase and frequency in the abstract and not other things like optical depth and condensed water path (for instance).

We have rephrased the abstract in two ways: the first to make it clear that we are not just referring to phase and frequency, and the second to clarify what we mean when we refer to incorrect and correct simulation of clouds. We hope that this better communicates our research findings and methods.

Line 2: 'The radiative bias, characterised by too much shortwave radiation reaching the surface, is attributed to the incorrect simulation of cloud properties, including frequency and phase. To identify cloud regimes important to the Southern Ocean, we use k-means cloud histogram clustering, applied to a satellite product and then fitted to nudged simulations of the latest generation ACCESS atmosphere model. We identify instances when the model correctly or incorrectly simulates the same cloud type as the satellite product for any point in time or space. We then evaluate the cloud and radiation biases in these instances. We find that when the ACCESS model correctly simulates the cloud type, cloud property and radiation biases of equivalent, or in some cases greater, magnitude remain compared to when cloud types are incorrectly simulated.'

L44: what ensemble is being referred to?

We have clarified that this study used an ensemble of selected CMIP6 models.

Line 45: 'Shallow-cumulus clouds are found to be consistently underestimated by a selection of CMIP6 models, ...'

L60: What aspect of Bodas-Salcedo 2014 demonstrates a need for consistent evaluation techniques?

This sentence was not intended to convey that Bodas-Salcedo et al. (2014) demonstrates a need for consistent evaluation, but rather that it demonstrates an instance where evaluation techniques have been used across studies, highlighting the effectiveness of consistent techniques. We removed this sentence.

L62: it is unclear what the first two sentences of this paragraph are referring to. What are climate-scale runs? Why wouldn't the synoptic meteorology be the same? I think what the authors are getting at *is the difference between coupled and AMIP runs. However, there is not any guarantee that the synoptic state will be the same across AMIP runs and the authors just discussed Field and Wood 2007 above, which composites on synoptic state- making it immaterial whether low pressure centers and other synoptic features are occurring in the same place.*

We have revised the first sentence of this paragraph to make clear that we are referring to free-running, long (climate-scale) simulations. We have revised the second sentence to make clear that often with these runs only monthly output is available (suiting the 'climate scale') and that the synoptic state (regardless of the temporal resolution of the output) cannot be considered the same as that observed due to the free-running nature of the models. The intention of these types of simulations is for climate analysis, where long-term averages can be considered directly comparable to observations. We have further revised this section to be clear that some of the simulations often used for theses analyses are in the AMIP style, with prescribed SSTs/sea ice, but this however still does not allow the synoptic states to be directly comparable to observations. The reviewer correctly notes that previous studies have gotten around this issue by using synoptic typing, when high-enough temporal resolution has been available. We point out that the synoptic typing done however often limits the analysis to particular conditions (eg. the predominant focus has been on cyclones, end even narrower, the cold sector of such) and ignores possible compensating errors in other synoptic situations. While cyclones are extremely common in the SO and may be the lead contributor of the radiative bias, other synoptic situations such as high pressure systems and frontal systems are also of worthy of examination. In this study, we wish to examine the entire system, without pre-conceived ideas of model performance (eg. limiting to one specific condition).

Line 62 : 'The majority of the aforementioned cloud-regime studies have compared free-running simulations, such as those performed for the CMIP experiments, including the Atmospheric Model Intercomparison Project (AMIP) where sea surface temperatures and sea ice concentrations are prescribed. However, with free-running simulations, the depth of analysis is limited as the synoptic scale meteorology cannot be considered the same. Some studies have used synoptic compositing to alleviate this issue, where certain synoptic situations can be compared like-for-like, and location and timing is then considered irrelevant. However, these studies are often limited to one synoptic type, such as cyclone centers, ignoring a number of other synoptic situations relevant for the SO, as well as any compensating errors that may exist. Additionally, focusing on just one synoptic condition follows a pre-conceived idea of the error, which may or may not hold true for newer model generations.'

L75: It is somewhat vague what the authors mean by 'incorrectly' or 'correctly' simulated... Is this just in terms of phase and frequency, or in terms of all characteristics in a more abstract way?

We have revised this sentence for clarity:

Line 81: 'By using a nudged simulation, we are able to composite and evaluate days and locations when cloud regimes are correctly (ie. are the same as what was seen by the satellite) identified by the model, as well as instances when the model incorrectly simulates the cloud regime. We aim to answer the following question: if the model simulates the correct cloud structure for the time and place, is the radiative bias improved?'

L175- It's a little ambiguous here whether the authors are referring to IWP and LWP averaged over the grid box, which is what the model outputs (aka clivi and clwvi-clivi), or if they are talking about in-cloud liquid and ice water path, which is what MODIS would see. It is also somewhat mysterious how propagating errors would affect LWP and IWP and not other cloud properties in COSP. Some additional discussion of this is needed to instill confidence in their evaluation.

We have used the grid box mean and performed the appropriate calculations to ensure the fields are comparable. We have made this clear in the text.

Line 186: 'For the LWP and IWP, we are considering the grid box mean (as opposed to the in-cloud mean).'

We have also expanded our discussion of propagating errors. Here we explicitly refer to the documented poorer retrieval of the cloud effective radii, which is used in the calculation of the LWP/IWP, hence the errors of this top-level retrieval are passed through to the LWP/IWP. We have chosen to use the raw model fields in this instance because we do not trust the retrievals of effective radii (which we also evaluated, but have not shown), and their impact on LWP/IWPs. We have added the text below to clarify this point and removed reference to propagating errors.

Line 187: 'IWP and LWP are reliant on R_{eff} retrievals, which as discussed above, are less well captured in satellite products.'

L193: Consistent with which previous studies?

We have removed this statement.

L235 and 245: Is CFL/CFI random overlap, or just what is seen from space? Could biases be driven mostly by this cirrus in the model if it is just what is seen from space, with minimal relevance for the PBL cloud that drives SWCRE?

As we understand, the CFI/CFL fields are the fraction of pixels successfully retrieved by MODIS, as seen from space, which the ACCESS-AM2 model replicates with the COSP simulator package. The point you make about cirrus is a good one, however, in the Southern Ocean, these cloud types are very infrequent (eg. see Mace et al. 2009 https://doi.org/10.1029/2007JD009755) so we do not expect them to be playing a large role in the SWCRE.

L249: Consider citing: Mülmenstädt, J., Salzmann, M., Kay, J. E., Zelinka, M. D., Ma, P.-L., Nam, C., et al. (2021). An underestimated negative cloud feedback from cloud lifetime changes. Nature Climate Change, 11(6), 508–513.

Field, P. R., & Heymsfield, A. J. (2015). Importance of snow to global precipitation. Geophysical Research Letters, 42(21), 9512–9520.

We thank the reviewer for bringing these papers to our attention. We have now removed reference to precipitation in this particular sentence, but we have included the Mulmenstadt work in our discussions.

Line 265: 'Too few liquid clouds which are instead simulated as ice clouds, will result in clouds that are more optically thin causing not enough short wave radiation to be reflected out to space.'

Line 611: 'Furthermore, in this work we have only considered a few cloud properties that impact the radiation budget. Other important factors, such as the number and size of cloud droplets, precipitation phase and amount, or other thermodynamical properties are likely to impact the absorption and scattering properties of clouds and their lifetime (for example: Mulmenstadt et al. 2021).'

L253: again, it is unclear if the authors are comparing in-cloud LWP and IWP to area- mean LWP and IWP.

Please see our response to the previous comment on the definition of LWP and IWP.

L273: This discussion is fairly qualitative in terms of relating various cloud properties to radiative bias. Quantitative estimates of how (for instance) cloud fraction relates to radiation exist:
Bender, F. A. M., Engström, A., Wood, R., & Charlson, R. J. (2017). Evaluation of Hemispheric Asymmetries in Marine Cloud Radiative Properties. Journal of Climate, 30(11), 4131–4147.
https://doi.org/10.1175/JCLI-D-16-0263.1
Can the authors show whether the CF bias in their simulations can explain the actual radiative bias?

We thank the reviewer for bringing this paper to our attention. We have added this reference to our conclusion section.

We agree that this paper is more qualitative in its analysis, though we do not believe this work is an exception to the literature norms in that respect. We currently have in preparation a follow up paper that takes a much more quantitative view on diagnosing the relationship between cloud properties (including cloud fraction) and radiative biases. We will submit this work shortly (pending this paper's acceptance).

Line 602: 'The work presented in this study has provided a relatively qualitative view of the cloud and radiative biases associated with cloud types. Studies such as Bender et al. (2017) have used more quantitative methods to evaluate the role cloud fractions play in determining the radiative balance. Specifically, Bender et al. (2017) examine the distributions of cloud albedo and the associated cloud fraction in CMIP5 models and their linear relationships. In further exploration of the data used in this work, we note that most of the relationships between cloud properties and radiative bias are in fact non-linear, which is also highlighted in this work by the differing relationships observed by latitude/cloud type. Similar results have been noted in Bodas-Salceado et al. (2016), where the LWP and radiative biases were not found to be as tightly coupled as expected. We will present findings of a quantitative analysis employing machine learning to understand the role cloud properties play in determining the radiative bias in an upcoming paper.':

Section 5.2.1: this section and the associated figure 7 are quite hard to follow. The authors may benefit from more clearly distinguishing errors in RFO and in cloud properties for a given cluster. The writing is somewhat ambiguous- clouds are referred to as being 'correctly' simulated- is this in terms of getting enough of them, or in terms of them looking right when they show up? In particular, the second paragraph of this section could be improved by using fewer vague qualifiers ('comparatively well captured', 'relatively well captured', 'somewhat correct',...) what is the baseline for these statements? These statements are then used to make causal statements about what biases in clouds are leading to biases in radiation, but without any support – wouldn't it be possible to do a more quantitative assessment of where biases in SWCRE are coming from?

We have revised the beginning of Section 5 to better describe our terms 'correct' and 'incorrect' and we have used this language more consistently throughout the text.

Line 389: 'One strength of comparing a daily, nudged, simulation to daily MODIS fields is the ability to make direct comparisons in time and space. As the synoptic meteorology is considered to be the same in the model and the observed conditions, we therefore expect that the model microphysics, if accurate, would generate the same cloud types that the large scale dynamics prescribes. With this assumption, we are able to isolate instances (points in time and space) where the model simulates the same cloud type as MODIS, which we define as 'correctly' simulating the cloud type, and the instances where the model simulates a different cloud type, which we define as an 'incorrect' cloud type assignment by the model. We demonstrate these definitions in Figure 6a and b. '

Figure 7 is pretty hard to follow. The authors may need a cartoon with annotations or something to illustrate this. There are dots, colors, outlines, months, clusters, and 5 different quantities. A single cartoon for one of the subplots would be helpful.

We have now included a diagram that helps to explain Figures 6, 7, 8 and 9. We have also tried to reduce the complexity of these figures by removing some of the information.

Overall, I would suggest moving the summary of the findings before section 5.2.1 to give the reader an overview of what is going to be discussed.

We have significantly revised Section 5 to improve clarity. We have also revised the introducing paragraphs so that the reader has a clearer idea of what to expect. For these reasons, we have not moved the summary to the top.

L555: the authors bring up a good point- is some of their RFO bias simply due to nudging? Can the k-means clustering be replicated on a free-running simulation to see what the biases look like?

A study examining the effect of nudging on cloud properties would be of interest, however is not within scope for our funding. We note that for cloud studies, nudging the temperature can change homogeneous ice nucleation rates, impacting cloud phase (Zhang et al. 2014, doi:10.5194/acp-14-8631-2014). This may influence other cloud properties, such as the optical depths and heights used in this work to develop the cloud types and subsequently the RFOs. However, Zhang et al. (2014) also showed that the largest impacts of nudging were felt in the tropics, including particular for convection. We speculate that due to the predominantly large-scale nature of the Southern Ocean (eg. fewer convectively driven clouds) and with temperatures already close to zero or below, the impacts of nudging will be less.