Author Response and Manuscript Revision

Sonya Fiddes

August 2025

Editor comments

I concur with all three reviewers that this is valuable work, addressing a key source of uncertainty in climate models. All three reviewers request clarifications to the text that may help readers better understand the methodology used, as well as the robustness and interpretation of findings. I agree with Reviewer 1 that it would be valuable to explain whether recent GLOMAP-mode aerosol scheme updates should be expected to impact model performance compared with the version analyzed here. Thank you in advance for your work to address reviewer comments.

Dear Dr. Ann Fridlind

Please find our response to the Reviewer's comments and our revised manuscript with an updated title: 'The ACCESS-AM2 climate model underestimates aerosol concentration in the Southern Ocean; improving aerosol representation could be problematic for the global energy balance'. We have addressed each comment in detail, including specific comments as to the model updates. We have spent considerable effort on clarifying the text and figures to make them more readable and easier to understand. I (Sonya - lead author) would also like to thank you, the Editor, the journal and the Reviewers for your patience and support while I was on parental leave.

Kind regards,

Dr. Sonya Fiddes

Response to Reviewer 1

We thank you for your in depth review of our paper. We have clarified the points you have raised with respect to both the model structure and language. Please see our detailed responses below.

Title and sub-titles:

- *The 2nd part of the title is not clearly supported by the results and should be changed to better reflect the paper content.*

We have updated our title to be more specific in line with the Reviewers comments.

'The ACCESS-AM2 climate model underestimates aerosol concentration in the Southern Ocean; improving aerosol representation could be problematic for the global energy balance'

- *Sub-title of section 3.4.2 is incorrect. The content describes data processing rather than statistical methods.*

We have updated this to 'Data Processing' as suggested

Vague language:

- *Language is imprecise in parts of the article, leaving the reader to guess the author's intended meaning. The discussion section stands out as particularly vague. One key example: in the first paragraph of section 6, the phrase 'as a whole' is unexplained, yet this seems to be the primary recommendation of the paper. The authors need to take the time and space to frame their hypothesised model development framework in more detail and with greater clarity to be convincing.*

We have made every effort to be more precise in our language. We have now revised the discussion, removing or explaining imprecise language, such as 'as a whole'. We have also more clearly laid out our argument and hope that the Reviewer may now be convinced as to its robustness.

For example:

Line 589: 'It also demonstrates the internal complexity of the aerosol population and the need to consider how each component influences the size and composition of the entire burden, rather than as individual (compositional) populations.'

We have also made our primary findings of the paper more clear in the conclusions:
Line 698:'We draw two main conclusions from this work, corresponding to two different regions of the Southern Ocean. We suggest that better capturing the biological influence on aerosol may lead to limited improvements in the aerosol-cloud-radiative system of the Southern Ocean's sea-ice regions, where the radiative bias is at its worst. We also show that in order to reduce the uncertainty of cloud feedbacks and the energy balance in the northern parts of the Southern Ocean, improving the aerosol alone is not effective (in fact is detrimental) but may be a pre-requisite for improving aerosol-cloud interactions due to the influence of aerosol composition and size.'

Other examples of vagueness in language that need to be addressed include:

- *In section 3.1, the description of the model indicates the ACCESS model was run in atmosphere-only mode. So, the model is essentially UM10.6 GA7.1 with GLOMAP-mode, using CMIP6 and CEDS emissions, nudged towards ERA5 data. The coupled aspect of the model seems irrelvant yet ACCESS is framed as the model being evaluated. What is ACCESS actually adding to the simulations?*

The Reviewer is correct that the model, run in atmosphere only mode, is based on UM10.6, GA7.1, however it also includes the Australian land-surface model CABLE, as opposed to the JULES model. This is stated on Line 130. The model, even when being run in atmosphere only mode, is referred to as ACCESS, with the second part of the name (eg. AM2) referring to the specific configuration. We make reference to the coupled model, as much of the documentation for the atmosphere only version is also captured by the documentation for the coupled model (eg. Bi et al. 2020).

- *Line 198 to 202: This meaning of this paragraph is hard to unravel. Clearer language is needed.*

We have revised this paragraph as follows to bring more clarity:

Line 221: 'We have also produced a daily, annually varying DMS dataset derived from output of the ocean component of ACCESS, ACCESS-OM2. We refer to this experiment as 'OM2 DMS'. Details of ACCESS-OM2 and the simulation used to produce the DMS output can be found in Kiss et al. (2020) and Sections 2.1 and 3.1 of Hayashida et al. (2021). ACCESS-OM2, in this case has used the atmospheric boundary conditions from the Control* experiment to drive the model (instead of reanalysis). We highlight that this DMS data set differs from the previous in that it is not a climatology - it is an annually varying dataset at a daily scale, able to respond to atmospheric and oceanic forcings. '

- *Line 304: GLOMAP-mode provides these values, though they may not have been selected by the authors. This description needs to clearly state this was a choice, rather than a model deficiency.*

It was not our intention to imply that this was a model deficiency. In our configuration, GLOMAP-mode outputs the N3, CCN50 and CCN70 by default, not CCN at a specific supersaturation. To more accurately represent the critical activation diameter for the Southern Ocean, we needed to calculate additional cut-offs, including CCN40, which are used in this work.

- *Figure 3 caption: 25th and 75th percentiles of what?*

The caption (Figure 3, 4 and 5) have now been updated to be more specific: for example:

Figure 3: '... For all subfigures, the 25th and 75th percentiles of the daily mean N10 are shown by the shaded range for the observations and control run. ...'

- *Line 340: Standard deviation of what? Calculated from which data?*

We have updated this to:

Line 383: 'The standard deviation of the daily mean N10 is also underestimated on average, where the control run for KGC has a mean standard deviation of $123\,\mathrm{cm}^{-3}$ compared to $342\,\mathrm{cm}^{-3}$ in the observations.'

We have also updated other sections of the text where it may be unclear as to which data we are referring to.

- *Line 463: The meaning of the final sentence is obscure and unreferenced.*

We have removed this sentence.

- *Constraint/constrained is used incorrectly in the article. I think the authors mean 'restricted to', 'in', or 'limited to'. Constraint has a specific meaning related to model uncertainty.*

We have updated two instances in the text in line with your suggestion.

- *These is extensive use of accronyms, which may be considered appropriate for some readers, but reduces readability. Particularly, readability is reduced by using accronyms for observation stations.*

We have removed reference to the SYO acronym (left in by mistake). We have also removed the acronym for Macquarie Island (MI). The acronyms for Kennaook/Cape Grim and ship-based campaigns have been retained.

Presentation of results with more obvious scientific reasonining:

Much of the manuscript needs to be rewritten to highlight key discoveries to the reader. The authors should consider where meaning is assumed and could be clarified. Additionally, the text often contains only statements about model-to-observation comparisons, without interpretation of meaning. Occassionally, statements conflict with results, which suggests they've not been considered deeply.

We have revised the Results section to be more clear and remove any statements of conflict. We note that we have a dedicated discussion section for more in depth interpretation of the results, which considers all of the results together rather than the individual as presented in Section 4. In line with a later comment by this Reviewer, we have tried to reduce the 'long-winded'-ness of this section but will keep the in-depth interpretation of meaning in the discussion section.

Some essential changes include:

- *Section 3.4. The first paragraph here is unneccessary. Nothing of value is added, so this should be removed.*

It has been removed

- *Line 335: First sentence is confusing.*

We have changed it to:

Line 379: 'Figure 3a-c shows the modelled and observed N10 concentration seasonal cycle for KCG, Macquarie Island and Syowa.'

- *Paragraph starting line 360: In this location, the NPF scheme test is the only sensitivity test to shift the model from biased low to biased high. Some interpretation of results is needed here. This result implies the persistent model bias might be partially overcome by implementing a more sophisticated NPF parametrization. Also, the simplicity of the NPF parametrization needs to be mentioned in this section to help the reader understand why the improvements are spatially restricted.*

We have included some discussion about the NPF results section. The Review raises a valid point that a more sophisticated NPF scheme may yield better results that can increase the small sized aerosol in a more realistic way. We have revised this paragraph to include the following:

Line 410: 'This large increase in small-sized aerosol may be a result of several factors, including the relative simplicity of the GLOMAP-mode BL NPF scheme (a binary scheme outlined in Section 3.2.1), the influence of terrestrial airmasses (which contain emissions of VOCs that mediate the BL NPF, despite our efforts to filter these influences out) or aerosol pre-cursors. A more complex NPF scheme, such as those discussed in the Introduction may yield more realistic results, while greater investigation into the observed and modelled aerosol and aerosol precursors is called for in the region.'

- *Section 4 has many long-winded descriptions that do not lead to insights or statements of how the results affect model interpretation.*

We have revised this Section to make it more concise. We note that much of our interpretation of the results is presented in Section 6.

- *Section 4.3: There is no mention of the BL NPF sensitivity test here, even though it is the only test to reduce activation ratios.*

Yes, good point. This was an attempt at being concise as we do not believe the BL NPF simulation to be realistic. But we can see that these results still deserve a mention as this Reviewer suggests. We have added the following:

Line 539: 'The BL NPF simulation in most cases, particularly for the northern regions, reduces the activation ratio, demonstrating its large production of small sized aerosol, which we suggest is unrealistic. '

- *Line 563: The meaning of the first sentence is incongruous with the results.*

We have clarified this sentence to read:

Line 611: 'We find that BL NPF had little impact on the regions of the Southern Ocean and Antarctic least influenced by terrestrial airmasses of the mid-latitudes (south of 45S). For regions where terrestrial airmasses are common (eg. northern latitudes and KCG), turning on BL NPF strongly overestimates small sized aerosol, which we suggest to be unrealistic. '

Missing detail and context:

As mentioned above, some sections are heavy with text, whilst others lack detail and critical information.

- *For example, where the Humphries data set is introduced, no context is provided for why it might be better, or more useful, than previous data sets. Furthermore, some sense of the motivation for including the specific sensitivity tests chosen would be extremely useful in the first paragraph of section 3.2.*

We have added some text around why the Humphries data set is used:

Line 247: 'Most of these observations have been described, collated, quality controlled, harmonised and evaluated in Humphries et al. (2023). The Humphries et al. (2023) paper provides the first seasonal and latitudinal description of Southern Ocean aerosol properties, providing an ideal basis from which to perform a modelling evaluation for this region.'

Additionally, we have now included some motivation for the sensitivity tests, as suggested by the Reviewer at the start of Section 3.2:

Line 170: 'These sensitivity tests range from realistic and established updates through to some experimental only changes. They include tests that bring the model in line with recent UM configurations (e.g. the inclusion of primary marine organics), the use of updated ancillary data (i.e. the new DMS climatology), and examining the applicability of existing parameterisations that are usually not used in this region (i.e. boundary layer new particle formation). Furthermore, changing the sea salt parameterisation and using a model-derived daily updating DMS field are more experimental but are useful for future model development.'

Other specific examples of missing detail/context include:

- *Line 207: Why are time-varying DMS datasets preferable? Need to say what value is added.*

We have included this information as suggested:

Line 228: 'The benefit of a daily varying dataset is that it is able to respond to atmospheric and oceanic forcings, such as sea surface temperatures or wind speed. This method can present, potentially, a more tightly coupled system, and if the parameterisation is accurate, yield more realistic DMS fields (including DMS in the water and the atmosphere). Some modelling groups are already adopting online DMS production (e.g. Bock et al. 2021), so this is a first step towards this goal for ACCESS. '

- *Line 222: 'underway' needs a description*

We have defined 'underway' as:

Line 256: ' 'underway' (automatic observations taken continuously while the ship is operating)'

- *Line 317: Why isn't the assumption made that the gridbox containing the observation would be the best comparison to make? There is no explaination for why the authors are even considering using a gridbox the the SW of the station.*

We have added a more detailed explanation of our choices re: grid box location.

Line 254: 'This is also true for KCG, where choosing a gridbox to the south-west of the station, as is normal practice for this location when studying baseline airmasses (which are not influenced by terrestrial air), resulted in poorer performance.'

- *Line379: The authors state 'This is a key area of development for GLOMAP-mode'. This statement needs to be put in context. The suggested model developments are only important if the priority is a model with increased skill at simulating aerosol concentrations over remote polar regions. The authors have assumed this is the case, with an implied further assumption that aerosol in these regions are more climatically important than aerosol elsewhere.*

This was not our intention. But to avoid confusion, we have removed this sentence.

- *Section 4.1: Some brief description of the overall under-prediction of CCN concentration and seasonal cycle amplitude, and what this implies should be given up front. Currently, this message is hidden amongst discussion of individual simulations.*

We have added the following:
Line 375: 'In this section we show that the ACCESS-AM2 model strongly underestimates both the N10 and CCN concentrations, which points to issues with the model's ability to accurately represent the aerosol population. We have limited our discussion in this section as the results will be analysed and interpreted together in Section 5. '

Old model version:

The authors have evaluated the impact of structural changes to model parametrizations, using a relatively old version of the GLOMAP-mode aerosol scheme, without reference to published model changes that would affect results. Results in this article need to be discussed with reference to latter model versions and with some consideration of how recent model developments may impact results presented here.

- *For example, no reference is made to the inclusion and evaluation of primary marine organic aerosol in later model versions. Additionally, sea salt density has been updated, as has deposition velocities via land surface representations, both of which would affect the sensitivity test results.*

Thank-you for highlighting these changes in the model. We note that on Line 689 we did make reference to the fact that 'Switching on PMO and re-scaling DMS brings ACCESS inline with more recent versions of the UM global atmosphere configurations' however we take the point that this should have been made more clear. We have now done this in the methods section as well.

Line 158: 'The DMS flux is scaled by a factor of 1.7 to take into account the lack of PMO, which are not switched on by default (Mulcahy et al. 2020). This is different in later versions of GLOMAP-mode, which returns DMS to a scaling of 1 and turns of PMO (Mulcahy et al. 2020).'

As for the changes to the sea salt density, we note that this change was adopted in GLOMAP-mode in ACCESS-CM2 (and therefore ACCESS-AM2), though this has not been documented. We will rectify this in the methods section.

Line 160: 'SSA emission fluxes are calculated using the wind-speed parameterization source function developed by Gong (2003), and include the updated sea salt density as per Mulcahy et al. (2020).'

We have not found evidence in the model branches/code that the updates to the deposition velocities were included in this version of ACCESS, so we have made comment as the Reviewer suggests about how this may impact our results in our Discussion section.

Line 656: 'A later update to the GLOMAP-mode dry deposition velocities has lead to increased coarse-mode deposition velocities that reportedly impact the sea-salt aerosol distribution (Mulcahy et al. 2020). We speculate that this update may reduce the magnitude of impact of our SSA changes on the CCN.'

Figures:

- *Font size in figures is sometimes too small. Additionally, thicker lines with better color contrast, or some other method, is needed to distinguish between simulations.*

We have increased the line thickness and font size in the figures. We have left the colours as is in order to ensure they are colour blind safe, and note that this is why we have also used different line styles.

- *On line 511, the authors state they have evaluted other cloud properties, which is essential to make a complete analysis of the impact of the sensitivity tests on aerosol, clouds and aerosol-cloud interactions. Equivalent figures should be included in a supplement, so the reader can interpret the wider effects themselves.*

We take the point from this Reviewer, and note that this is something we discussed extensively amongst the authors. We also note the opposing view from Reviewer 2, to remove the LWP discussion as it is not evaluated against observations. In this light, we have decided to leave the text as it is for two reasons: a) the cloud-radiative biases in ACCESS-AM2 have been well documented in our previous work, and b) we do not want to take away from the main point of this study, which is the aerosol evaluation. We acknowledge that we have made some strong statements (which this Reviewer disagrees with) about the climatic impacts of the changes to the aerosol scheme. We have lessened these statements to better reflect what our results are, and to highlight them more clearly.

References:

Some additional references that have been overlooked include:

- *Schutgens et al. (2017) doi.org/10.5194/acp-17-9761-2017 in section 3.4.1*
-

We thank the Reviewer for bringing this work to our attention, we have now included it in the methods section:

Line 364: 'Schutgens et al. (2017) recommend collocating model and observational data at hourly intervals to reduce representation error. However, for these simulations this approach was not feasible. '

- *Additional literature evaluating more sophisticated NPF schemes and the climatic effects of those schemes.*

We note that more sophisticated NPF schemes are briefly mentioned in the Introduction. We have now also included the following in our Discussion section:

Line 615: 'Most global atmosphere models use classical nucleation theory involving binary NPF, however, more complex ternary or ion nucleation parameterisations have also been developed. For the Antarctic, ion nucleation of sulfuric acid with ammonia (sourced from sea bird colonies) has been suggested to be an important pathway for nucleation (Lee et al. 2019), implying that a more complex NPF scheme could benefit this region. However, significant updates to the chemistry in ACCESS would be required to include such sources.'

- *The final sentence in section 4.2*

This sentence has been removed.

- *Model structural changes implemented after this model version, particularly where they may affect interpretation of results here (e.g. https://gmd.copernicus.org/articles/13/6383/2020/)*

We have addressed the structural changes as per the previous comment in Section 3.1.1.

Spelling and syntax:

- *Line 101: '(GLOMAP)' over-used*

  We have rephrased this sentence as:

  Line 113: '... which includes the Global Model of Aerosol Processes (GLOMAP)-mode aerosol scheme... '

- *Line 149: 'volcinic-sourced'*

  Hyphen added

- *Equation 7: Numerator should be 'CHL'*

  Numerator fixed

- *Line 261: Remove '/,'*

  LaTeX syntax fixed

- *Line 290: SYO not defined*

  We have removed all reference to SYO throughout text

- *Line 310: inline*

  All instances of 'inline' have been change to 'in line'

- *Section 3.4.2: This could easily be a single paragraph*

  Has been made one paragraph

- *Line 353: missing 'is'*

  Fixed

- *Line 605: bis*

Fixed

Response to Reviewer 2

We would like to thank this Reviewer for their input and appreciate their careful comments. We have addressed each of their comments below.

Major issue 1: additional context needed on significance of model-observation differences

- *At present, the manuscript text presents a large number of percentage differences between observed and simulated SO aerosol properties. The only context for the significance of these properties comes from the figure shading around 25th-75th percentiles for the control simulation (not of the other 7 simulations) and observations. From only this information, it is very challenging for the reader to evaluate whether these percentage differences are significant and to assess if the first part of the manuscript title ("The ACCESS-AM2 climate model strongly underestimates aerosol concentration in the Southern Ocean") is justified. I outline below a few different sources of uncertainty that are missing or only mentioned qualitatively in the present text.*

We thank the Reviewer for their concern in this respect. We agree that more context would be desirable for this work, however, adding this information comes at a cost of many more figures and tables, or increasing complexity of figures (hence only showing the percentiles for the observations and control).

We have now included on Figures 2-4 the 25th-75th percentiles for each of the voyage data points, and the annual means for the station data. We have also added this range for the PMO+H22 simulations seasonal cycle. We hope that this provides more context.

- *Point measurements vs grid cell means: The manuscript does not appear to address at all that it is comparing observed point (or near point when integration time is a few minutes) measurements to grid cell mean simulated measurements. This could induce large biases depending on the shape of the sub-grid scale distribution of aerosols, especially for the coastal sites. Some context on the sub-grid scale distribution could be gained from plotting distributions of measurements near in time and space and assessing skewness.*

We have considered this problem, which as the Reviewer suggests could induce biases especially near coastal sites. As stated on line 324, we linearly interpolated the daily mean ship location to the model grid. For the open ocean and a moving platform, we have deemed this satisfactory. For the land-based measurements, Macquarie Island, is not resolved by the model's land fraction, so any coastal interaction will not be resolved by default.

For Kennaook/Cape Grim, the standard practice in the Australian community doing model comparisons to the station is to take an oceanic grid box upwind from the station in order to account for the coastal interference in the model, which is usually removed from the observations via the baseline filtering (as done in this work). This method has not been published and remains to be tested vigorously. Nevertheless, we did perform this analysis, but found that the performance of the model was worse. This is discussed on line 316 onwards.

At Syowa, terrestrial influence is not of a concern for this work (but is of future interest) and so no filtering was applied to isolate oceanic only airmasses, and no adjustments to the grid boxes were made.

This is in all to say, comparing point based observations to a coarse modelled product is never bulletproof, but we are satisfied that our methods can provide results that are representative of the modelled biases. Assessing skewness based on nearby gridboxes would not yield information about sub-grid scale distribution, as this latter is strongly influenced by processes (e.g. convection) that are not explicitly resolved in our simulations. Lack of resolution of finer scale processes is a necessary compromise of global climate models.

- *Instrument and instrument simulator uncertainty: The manuscript is missing context on instrument uncertainty. Further, for the CCN comparison, the manuscript should provide uncertainty estimates due to interchanging a super-saturation-based threshold for the observations and a size-based threshold for the model. This translation will be uncertain due to assumptions needed to translate particle activation supersaturation to particle activation size (as is somewhat mentioned* *presently but not made quantitative) and due to imprecision in supersaturation control in the instruments. Ideally this uncertainty could be propagated through to the reported model results, but if this is too difficult, sensitivity tests of different radius thresholds should be reported. This latter test seems straight forward from the code in process_aer_along_ship.ipynb. Additionally, some of campaigns (e.g., MICRE) have both 0.5% supersaturation aerosol count and aerosol size distribution measurements. A comparison of the two CCN methods can be made directly on the observations and its error assessed.*

We have not included a full description of each instruments uncertainty/sensitivity in this work as they are well documented in the prior papers (eg. Humphries et al 2021a). We will make this more clear in Section 3.3. Because the measurements were taken on different instruments at different times, propagating a consistent uncertainty for all the observations is difficult.

Line 247: 'Most of these observations have been described, collated, quality controlled, harmonised and published in Humphries et al. (2023), providing the first seasonal and latitudinal description of Southern Ocean aerosol properties. The Humphries et al. (2023) paper provides the ideal basis from which to perform a modelling evaluation. '

For the CCN comparison, yes, uncertainty is unavoidable when translating between a supersaturation and a size-based threshold. We have now included more text around this in Section 3.4.1.

Line 336: 'The model does not provide CCN diagnostics at a specific supersaturation, but provides CCN at selected dry diameters. The modelled size distribution can be used to calculate CCN at any particular activation diameters.'

Line 347: 'The calculation for estimating a critical diameter for a given supersaturation is imperfect and does increase the uncertainty of our results. We have also tested the critical diameter at CCN50 (which more closely matches observed aerosol populations at 0.5% super saturation (Fossum et al. 2018), which showed marginally better results near Kennaook/Cape Grim (no longer overestimated), and marginally worse results elsewhere. However, our critical diameter calculations were consistent in reporting 40nm as the cut-off for the modelled aerosol size distribution. '

The Reviewer is correct that aerosol size distribution measurements are published for some of the voyages (eg. Alroe et al. 2020 https://doi.org/10.5194/acp-20-8047-2020). In fact size distributions are available for most of the voyages, but still require processing. This is an area of high priority for us, and we currently have a student working on harmonising all size distribution data for all Australian SO voyages, providing over 10 years of data. For this reason we have not included the preliminary analysis we did on this here as we hope to perform an extensive evaluation in the near future.

- *Variability of non-control model runs: The manuscript should comment in some form on the variability (25th-75th percentile) of the non-control simulations. Ideally this could be visualized in Figures 3-4 (though this might be too visually confusing) or the supplemental.*

The reviewer is correct that we have not shown the percentiles for the sensitivity simulations as they made the figures unreadable. We have now added them for the voyage data and for the annual station data. For the seasonal cycles, we have added just the PMO+H22 simulation (to that of the observations and control) to retain readability. We note that the range of results can also be inferred from the standard deviation which is also reported in the text.

- *Exhaust corrections: Some context on the magnitude of these corrections on the observed N10 and CCN at different sites/ships should be given as well as the uncertainty introduced by these corrections.*

I would hesitate to call the exhaust filtering a correction. The authors do not change any of the data, rather than eliminate data the appears to be contaminated by ship exhaust, which by nature, will not be found in a model. To be representative, we also eliminate those days from the model data. This approach does reduce our sample size, in some campaigns by over 75%, but this is why a large scale study such as this, using multiple campaigns is essential to build up the statistics. Much more detail about the filtering can be found in Humphries et al. 2019.

Specific comments

- *Abstract lines 13-14: "Our results indicate significant problems in the model's microphysical processes and with over tuning." Given how the manuscript is currently written, I believe this sentence should be rephrased to reflect that this is the authors' opinion. This statement is one possible interpretation of the manuscript's results. I see only one paragraph of the present manuscript referencing the tuning issue and only tangentially. (See Lines 606-611 for more detail.)*

We have revised end of the abstract to read:

Line 13: ' This significantly improved CCN in the marine regions, but resulted in detrimental impacts on the region's radiation budget, indicating that drastically improving the Southern Ocean's CCN budget may lead to poorer outcomes for the global climate.'

- *Abstract line 14: "We suggest this needs to be addressed in a holistic way." I only see part of one sentence explicitly supporting this abstract statement: "... [this] points to a need to consider model development in this space as an entire system rather than individual components." I think this should have slightly more discussion to be mentioned in the abstract.*

We have removed this sentence.

- *Lines 105-106: "By performing these evaluations [of N10 and CCN], the model biases associated with aerosol-cloud-radiation interactions around the Southern Ocean and Antarctic can be better understood and the degree of uncertainty reduced." Another sentence or two connecting why these aerosol properties specifically (especially N10) would accomplish this would make the introduction a lot stronger. (Just a suggestion.)*

We have edited this sentence to now read:

Line 117: 'By performing these evaluations of N10 and CCN, we can gain a better understanding of the modelled aerosol population and it's biases. Examining the population at two different sizes can give us insight as to how different species may be impact both the overall aerosol population as well as that of cloud-relevant size and the growth that occurs to this size. With this knowledge, we can outline how the model biases associated with aerosol-cloud-radiation interactions around the Southern Ocean and Antarctic can be better understood and the degree of uncertainty reduced.'

- *Section 3.2: The current model configurations tested (Table 1) all focus on aerosol sources. What about aerosol sinks? Even if it is out of scope to test these components of the aerosol scheme, too, the text should at least mention how the model represents aerosol sinks (e.g., coagulation, rain out) and discuss how these processes' parameterization might contribute to model-observation discrepancies.*

This is a good point and something that we have discussed within our group. On line 134 we do mention these processes within the model - coagulation, deposition, cloud processing, however these processes are not discussed in terms of their impact on the aerosol.

We have now added a paragraph in the introduction which discusses these processes, however, as the Reviewer notes, testing these parameterisations is out of scope within this work.

Line 91: 'Aerosol sinks, and how they are modelled, are also a key source of uncertainty. Aerosol can be removed from the atmosphere via dry deposition or wet deposition. Dry deposition is difficult to measure and evaluate. Regayre et al. (2020), after applying Southern Ocean observational constraints to a perturbed parameter ensemble, find that it is likely that a scaling factor for the accumulation mode dry deposition velocity in the Unified Model needs to be lower than the default value. Other observational studies have indicated that wet deposition (rain after coalescence of cloud droplets) is an important control of CCN variability in the Southern Ocean, particularly in relation to shallow convection

(Alinejadtabrizi et al. 2024) and stratocumulus (Kang et al. 2022). Given the tendency for models to produce too much light rain (Stephans et al. 2010), it has been suggested that wet deposition may be overestimated in models (Kang et al. 2022).'

- *Lines 299-301: "Some evaluation of the model['']s meteorology has been carried out, but is not shown in this work. It was found to be satisfactory, which is in line with our expectations due to nudging." Can this be presented in a supplemental? With such minimal description, it is not clear what is "satisfactory."*

We have not presented this work for the sake of brevity. Overall the model performed well which was expected given the nudging. We have now removed reference to this work so that we do not confuse readers and to not take away from the main focus of this paper.

- *Line 316-317: "At KCG, we have used the exact model gridbox that the station is located in, as choosing a gridbox to the south- west of the station resulted in poorer performance." This sentence seems to suggest at other locations the exact model gridbox was not used? If this is true, this conflicts with my understanding of Section 3.4.2 for a stationary location.*

We have edited this sentence to clarify that we are using the exact model grid box everywhere.

Line 353: ' The linearly interpolated model gridbox for each location (see Section 3.4.2) was used to perform the comparisons. This is also true for, as choosing a gridbox to the south-west of the station, as is often standard practice for this location, resulted in poorer performance.'

- *Section 3.4.2: Do ships cross multiple grid cells in a day? (Is the daily average lat-lon location problematic?)*

While we have not calculated the exact change in position for each ship each day to say with certainty, we would say that it is highly likely that a ship can cross more than one grid box in one day. This is especially the case for the resupply voyages where the ship may be travelling at full speed (approximately 12 nautical miles an hour) for up to a week to get to the Antarctic bases. A ship doing this can cover more than 150km in one day - the approximate size of a grid box. We have used a linear interpolation of the point location to extract the model data to account for instances when the average lat lon location is at the edge of a grid box.

- *Figures 3, 4: It is unclear exactly what "the monthly and annual median concentrations of N10/CCN" in the captions means, especially when compared to the methods description. Does this mean "monthly and annual median of mean daily concentrations of N10/CCN" or "median of monthly and annual concentrations of N10/CCN"? I assume the former from the methods, but this is not clear from the text. I am not sure whether the colors used are colorblind safe.*

You are correct that is is the former and we have made this more clear in the captions. We have passed all our figures through a colour blindness simulator to make the best efforts at presenting figures that are colour blind safe. This is also why we have used different line styles to aid visual identification.
Figure 3 caption: 'The monthly and annual median concentrations of daily mean N10 for at...'

- *Figures 6, 8: It would be helpful to compare the different model configurations' TOA SW radiative flux directly to the CERES product instead of plotting line contours where the control bias changes signs. Even just adding one panel with the zonal means would be helpful to support the manuscript's conclusions about which configurations improve the model relative to observations. I find it very difficult to get this important information out of the current plots. (Just a suggestion.)*

We have included Table 2 for to provide a quantitive view of the model changes compared to CERES. For this reason we won't add an additional panel.

- *Figure 7: Since LWP isn't compared to an observational product and is only mentioned as a key controller of TOA SW flux (already shown), this figure and its associated text don't seem necessary to the main manuscript text and could go in a supplement.*

Yes, this is something we discussed in depth as to whether we include or not, as I for one did not want this part of the paper to take away from the main findings of the aerosol evaluation (and also that we already have several papers published looking at the LWP biases). In the end we compromised and left this plot, but didn't go as far as to show the other cloud properties. We note that Reviewer 1 has argued for further cloud property plots, so in compromise again, we will leave the plots as they are.

- *Lines 511-512: "We note that we have evaluated other cloud properties, and the aerosol direct effect via clear sky radiation, but for brevity will not discuss them here." This seems unnecessary to mention if these evaluations aren't even roughly summarized and aren't included in a supplemental.*

We have removed this sentence.

- *Lines 606-611: This paragraph seems to motivate the second half of the title and the end of the abstract. It is not clear to me that these are unique conclusions from the methods presented and analysis shown. There is some discussion of why the experiments are probing what they are probing, but their implementations still contain such large uncertainties that it is unclear if scattered improvements in the simulated-observed CCN comparison demonstrates "improvement of the physical representation" of aerosols. This is why I suggest toning down the abstract above. Since the title includes "could," I think it is acceptably couched.*

Thank-you for this feedback. We have altered the abstract as you suggested. We have also edited the abstract to not overstate our results, removing this specific line.

Technical comments

- *Title: The antecedent for "it" is not clear from the title alone. The antecedent could be either "ACCESS-AM2 climate model" or "aerosol concentrations."*

We have changed the title to 'The ACCESS-AM2 climate model strongly underestimates aerosol concentration in the Southern Ocean; improving the aerosol scheme could be problematic for the modelled climate system'

- *All of the manuscript's equations are presented at the end of paragraphs and are not interwoven into the text.*

We will leave it to the copy editors to decide on final placement of equations.

Response to Reviewer 3

We would like to thank this Reviewer for their positive and constructive comments. We have addressed each of them below:

General comments

Section 3.2.3

- *I found Eqn 5 a little difficult to make sense of. The sentences "We note that the wind speed function is used here to represent surface tension of the sea surface microlayer (surface accumulation of organics). Higher wind speeds break this layer up, resulting in fewer organics being lofted into the atmosphere." are useful, but could some extra information be added to explain how the organic mass fraction responds to ocean chlorophyll-a and sea spray particle diameter.*

We have now included the following text about how the particle size and chlorophyll-a modulates the primary marine organic emissions.

Line 202:'Primary marine organic emissions are positively correlated to the seasonal cycle of CHL, acting as a proxy for biological productivity. The organic fraction of SSA is inversely related to the SSA particle size at sub-micron scales (the smaller the particle, the more organic fraction), while at super-micron sizes, the organic fraction is small and relativity constant.'

Figure 1

- *It is useful to see the climatologies plotted, however, it is not easy to see the differences between them. I appreciate that log scales are necessary, but would the authors consider using a more differentiated colour scale. Or perhaps using one of the data sets (maybe Kettle) as the "reference" and plotting the others as differences relative to the reference.*

We have altered to colour scale to try to make the colours less dark. We note that there is substantial literature on the differences between DMS climatologies, hence we have not made this comparison. We have added some more references to this effect in the text.

Line 216: 'Significant literature exists around the production and differences of DMS climatologies and we refer readers to these: (Hulswar et al. 2022, Lana et al. 2011, Zhou et al. 2024). '

Section 3.2.4

- *Could the authors please clarify whether the OM2 DMS simulation used the daily DMS values derived from the OM2 simulations or if a monthly mean of those daily values was used.*

We have rephrased this section to be more clear:

Line 226: 'We highlight that this DMS data set differs from the previous in that it is not a climatology - it is an annually varying dataset with daily resolution (we have changed the model to update DMS daily instead of monthly).'

Section 3.4, L298-300

- *Is there a reference, i.e. another piece of literature (not necessarily part of this work), where the performance of meteorology in the nudged model was assessed?*

We point this reviewer to two papers:

Uhe, P., & Thatcher, M. (2015). A spectral nudging method for the ACCESS1.3 atmospheric model. Geoscientific Model Development, 8(6), 1645–1658. https://doi.org/10.5194/GMD-8-1645-2015

Telford, P. J., Braesicke, P., Morgenstern, O., & Pyle, J. A. (2008). Atmospheric Chemistry and Physics Technical Note: Description and assessment of a nudged version of the new dynamics Unified Model. Atmos. Chem. Phys, 8, 1701–1712. www.atmos-chem-phys.net/8/1701/2008/

We have now included this in the text:

Line 360: '...given the model is nudged to ERA5, we expect the large scale flow to be accurate (Uhe & Thatcher 2015, Telford et al. 2008).'

Section 3.4.1, L319-322

- *"We also recognise that we have not performed a similar baseline filtering to the model (in part due to lack of radon in the model), but have applied the same baseline filtering to the model as what was developed for the observations." The above is not clear to me. I don't follow what the baseline filtering is, or the role Radon is playing.*

We apologise for the confusion here. We have now more clearly defined what is meant by 'baseline' and how radon is considered in defining this in Section 3.3.4 and referenced this in Section 3.4.1.

Line 210: 'The data presented here is the baseline filtered data (as described in Gras & Keywood, 2017). Baseline air is considered the worlds cleanest air, as unaltered by human activity as physically possible. At KCG, baseline air is identified as air that has come from the Southern Ocean where the wind direction was between 190 and 280 ° and the radon concentration (a marker of terrestrial influence) is below 100 mBq.'

Line 357: 'We also recognise that at KCG we have not performed a similar baseline filtering to the model data, in part due to lack of radon in the model (see Section 3.3.4 for details). Instead we we have matched the model data to the available daily mean baseline-filtered observations. '

Figures 3, 4, 5

- *Is there a reason for the offset horizontal lines in the subfigures? For example, in Figure 3a annual N10 is shown as horizontally offset coloured lines for the model experiments on the right hand side of the plot. Similarly in Figures d-g the lines representing seasonal mean N10 for the different model experiments are horizontally offset.*

Yes, we offset them just to make identifying different lines slightly easier, as some of them are very close together.

Section 4.1, L342

- *I don't think it's clear from the available data that the model mis-represents the seasonal minima at MI in May. The winter minimum N10 at MI appears to be more variable than at kennaook and Syowa and the model does not capture this. However, due to limited number of years of observations and model grid cell to point observation comparison, I think that it is difficult to conclude that the model mis-represents a seasonal minimum at MI in May.*

Yes, this is a good point. We have now altered the text as follows:

Line 387: 'The model's seasonal cycle is flat compared to the observations, indicating both missing sources of aerosol and missing seasonal processes. The control run does not capture the seasonal minima, which in the observations is shown in May after a steep decline through autumn, whilst for the model is shown in June (Figure 3b). Note that there are limited observations (only two seasonal cycles) and that there is greater observed variability (as shown by the shading) during winter. '

Section 4.1, L345

- *I'd argue that the model and observations both show less variation in winter. The values (for N10 etc) are smaller in winter compared with summer for both model and observations, so the variation as a fractional or percentage might better show if the variation was really much smaller in the observations.*

Yes, the text reads: 'The model again shows little variance in the winter periods, with larger variance in the summer.' which we think is in agreement with the Reviewers comment.

Section 4.1, L349

- *I agree that at Syowa the model does seem to simulate the minima too late compared with the obs. However, the model minima looks to extend from Jun-Aug, while the observed minima looks to be June (rather than May as stated).*

We have corrected this.

Line 394: 'Syowa has a minimum in June that is not captured by the model, which simulates the minima in August, although is generally low from May-August.'

Section 4.1, L365-367

- *Does ACCESS use VOC emissions ancillary files or calculate VOC emissions online? Can the authors say if there were there large VOC emissions over the first grid box? If there weren't large land-based VOC emissions over the first grid box, this might strengthen the argument that there are issues with the marine biogenics. Although I take the author's point that shifting the grid box did not help.*

ACCESS does use ancillary files for terrestrial VOC emissions (monoterpenes) (as per the CMIP protocol). The emissions over the grid box that KCG is located are small (by orders of magnitude) compared to other densely forested regions (e.g. the tropical and subtropical forests) however, they are not zero. In the text we have now re-iterated that we have attempted to filter out the days in which the terrestrial airmasses are experienced at KCG, but noting that removing this terrestrial influence all together from the model is difficult. We have ambitions to study these artefacts more fully with dedicated simulations beyond what we can accomplish with the existing simulations. We note that this section has changed somewhat due to other Reviewer comments.

Line 410: 'This large increase in small-sized aerosol maybe a result of several factors, including the relative simplicity of the GLOMAP-mode BL NPF scheme (a binary scheme outlined in Section 3.2.1), the influence of terrestrial airmasses (which contain emissions of VOCs that mediate the BL NPF, despite our efforts to filter these influences out) or aerosol pre-cursors. A more complex NPF scheme, such as those discussed in the Introduction may yield more realistic results, while greater investigation into the observed and modelled aerosol and aerosol precursors is called for in the region.'

Section 4.1, L380-384

- *Plots of aerosol size distribution would help diagnose how the model (re)distributes aerosol in the simulations.*

Yes we agree on this front. We have chosen not to show these results for two reasons. Firstly that we have only a few observed size distributions currently available to compare to, and second, our team is carrying out a body of work to process, quality control and harmonise 10+ years of observed size distributions, including for all the voyages used in this work. We are planning a significant body of work once these observations are published.

Section 4.2, L423

- *"... which could indeed be driven by sea spray, long range transport of aerosol" =¿ Add "or" before "long range transport of aerosol"*

Done

Section 4.2, L430-437

- *I agree that turning on the BL NPF has a small effect, but also worth noting that it's the only experiment that reduces CCN in MI and Syowa.*

The BL NPF simulation increases aerosol very marginally at MI and Syowa (the Reviewer may have gotten the lines mixed up (light red vs dark red, which we hope we have now rectified with clearer plots).

Section 4.2, L438-437

- *The contours in Figs 6-8 are quite hard to see. Could the authors consider changing the colour of these? Something like cyan might stand out more.*

We have tested out a range of colour combinations and the results are quite visually jarring. For this reason, we have retained the current colour ways, but have made the contour lines thicker.

References

- *Please update Bhatti et al., 2023 to the final version of the accepted maniscript.*

Done

Other points

- I agree that missing sources and mis-represented microphysics (probably aerosol and cloud) are large contributors to bias in cloud and RF over the Southern Ocean. Is there any reason to suspect loss processes (deposition) might be overestimated? Can the authors comment on how 'missing' marine VOC sources of VOCs (and possibly secondary organic aerosol, e.g. https://doi.org/10.1016/j.scitotenv.2021.145054) could affect clouds and climate over the Southern Ocean?

Yes, we agree with the Reviewer in that loss processes should also be examined. We have added some text around this in the Introduction.

Line 91: 'Aerosol sinks, and how they are modelled, are also a key source of uncertainty. Aerosol can be removed from the atmosphere via dry deposition or wet deposition. Dry deposition is difficult to measure and evaluate, however Regayre et al. (2020), after applying Southern Ocean observational constraints to a perturbed parameter ensemble, find that it is likely that a scaling factor for the accumulation mode dry deposition velocity in the Unified Model needs to be lower than the default value. This would result in a reduced sink of aerosol. Other observational studies have indicated that wet deposition (rain after coalescence of cloud droplets) is an important control of CCN variability in the Southern Ocean, particularly in relation to shallow convection (Alinejadtabrizi et al. 2024) and stratocumulus (Kang et al. 2024). Given the tendency for models to produce too much light rain (Stephens et al. 2010), it has been suggested that wet deposition may be overestimated in models (Kang et al. 2024). '

We have also added some comments about how marine VOCs may influence the Southern Ocean in our introduction:

Line 85: 'Additionally, marine volatile organic compounds (VOCs), such as isoprene, can reduce the atmospheric oxidative capacity by reacting with OH (as well as O_3 and NO_3 to a lesser degree) in the troposphere. Such VOCs, can also yield secondary organic aerosol and provide condensational mass, further influencing the clouds and climate. In the case of marine isoprene, this occurs on a much smaller scale than that of DMS (Yu et al. 2021), with isoprene concentrations being very low outside of phytoplankton blooms and biologically active coastal regions of the Southern Ocean Ferracci et al. (2021).'

- To help visualise the percentage changes in N10 etc across the model experiments the authors could consider including a matrix/table of percentage change colour coded to show an increase or decrease for the parameter (e.g. N10).

We appreciate the Reviewers suggestion here and have considered this carefully. Tables or matrixes that encapsulate all of our results are, bluntly, enormous. For this reason, we have pointed interested readers to our github page where this information can be found, but have not included it specifically here.

Line 367: 'Quantitative summaries of all our results can be found in the published code linked to this paper (see Code and Data Availability).':

Technical comments

- *Introduction, L22:*
- *Change "Aerosol affect. . . ." to "Aerosol affects. . . . ."*

Updated

- *Section 3.1.1, L141; Section 3.4.1, Line 310:*
- *A grammatical point, but I felt that "inline" should be "in line" in the text.*

We have updated all instances of this to be 'in line' with your suggestion

- *Section 3.3.1, L221: Not sure what 'underway' means here.*

We have defined 'underway' as 'automatic observations taken continuously while the ship is operating' and included this in the text

- *Section 3.3.3, L221: Change CO2 to CO2*

Done

- *Section 3.3.4, Heading and L265 (and Figure 3, 4, 5): I'm guessing that "kennaook" does use a lower case k, but in that case the legend in Fig 2 is the odd one out.*

We have updated these instances to be consistent.

- *Figure 3, 4, captions: "The monthly and annual median concentrations of N10 for at. . . " - Delete "for" or "at"*

Fixed

- *Fig 3-5, Captions L2: "For all, the 25th and 75th percentiles. . . " - Suggest adding "For all subfigures. . . " for clarity.*

Changed

- *Section 4.1, L360: I suggest a new sub-section here to report model simulations.*

We have added this for each of the Results (section 4) subsections.

- *Section 4.1: "For DJF, the BL NPF simulation is now overestimates the observations by 33%" - Please correct this.*

Fixed

- *Section 5, L518: Three best simulations rather than 4?*

Fixed

- *Section 5, Figure 7: Units on colour bar are g m-2. Units in caption are kg m-2.*

Fixed

- *Section 7, L639: Please change earth to Earth*

This sentence has been removed.