Dear reviewers,

Thank you for your thorough and helpful feedback. Based on your comments, we are pleased to submit a revised version of our manuscript ‘Assessment of an updated polar stratospheric cloud parameterisation for the UK Earth System Model (UKESM1.1) within the UK Met Office Unified Model (v13.9) using CALIOP and MLS observations’.

To address the comments, we have implemented the following main changes:

- Figures 1 and 2 have been amended to have linear colour bars and threshold lines that are calculated for each panel individually.
- Additional analysis has been included in Section 3.2 that compares the occurrence frequency of each PSC type in Antarctic winter months over the 2007-2011 period between CONTROL and $\mathrm{RUN}_{\mathrm{HH}}$.
- Denitrification for the 2009/2010 Arctic winter and dehydration for both the 2009/2010 Arctic and 2008 Antarctic winter have been included in Section 3.3 and the Supplementary Material.
- The Summary Section has been revised to clearly address the UKESM1.1 HNO_3 low-bias and inability to produce the range of optical properties in the CALIOP-defined ice region.

Below is our point-by-point response to all three reviewers’ comments. Our response is in blue and new text in the manuscript indicated **bold** and *italicised* text**.** The response to each reviewer comments can be found:

- Reviewer 1: Page 2
- Reviewer 2: Page 11
- Reviewer 3: Page 17

We have made other small changes to the text to improve the readability of the manuscript. These are all highlighted in the marked up revised manuscript.

The feedback from the reviewers has greatly benefited the manuscript and we believe it is much improved since implementing the suggested revisions. We would like to thank the reviewers’ and editor for all their time and effort and hope the revised manuscript meets the journal’s standards.

Sincerely,

Isabelle Sangha, on behalf of all the coauthors

Referee #1

This paper is well written with thorough citations, documenting all the before-and-after details of the model development, explaining all the methods used for data analysis especially trying to have apple-to-apple comparison with satellite data.

However, the results part still need some attention. The current SAD analysis is not convincing; there's missing necessary plots and analysis for Arctic and Water vapor. Once these two issures are address, more discussion need to be added in the summary session. I provide several major concerns here for your consideration.

1. Line 317-318: This statement is not correct. NATmix is always the dominant composition category in the CALIPSO Antarctic data (Pitts 2008). From Figure 3, it's hard to know if your new simulation of PSC categories are consistent with what CALIPSO shows for STS and NAT. CALIPSO shows NATmix usually dominant the mid-winter, while the STS dominant the early winter. you should show three panels of absolute values of STS, NAT and ice from RUN_HH. You ice might be ok, maybe a little bit too long into September. Your RUN_HH is not showing early winter SAD (later may to June), which should be dominant by STS. Also why does you simulated PSCs so low in altitude in RUN_HH? The CONTROL has better altitude representation (match the temperature threshold better and more consistent with CALIPSO cloud area). Is it because your background HNO3 is too low? What about your background water vapor? From your SAD plot, I am not convinced your new PSC scheme is better than before.

Reply: We thank the reviewer for this detailed and helpful comment and address each point in sequence.

Firstly, we would like to clarify that the statement at Lines 317 – 318 is consistent with Pitts et al. (2018), which we reference in support. Specifically, Pitts et al. (2018) (Fig. 16 therein) shows that STS is the most prevalent composition above 20 km in the early Antarctic winter, transitioning to NATmix dominance through mid-winter, before STS becomes prevalent again late in the season at lower altitudes. We therefore respectfully disagree with the referee on this comment and choose to leave this statement unchanged.

Secondly, following the reviewer's helpful suggestion, Figure 3 (now Figure 5 in the revised manuscript) has been updated to include three panels showing the absolute zonal mean SAD of STS, NAT, and ice from $\mathrm{RUN_{HH}}$ alongside the existing difference panels ($\mathrm{RUN_{HH}}$ – CONTROL). The updated figure is included below (Fig. 1). The inclusion of the absolute mean SAD from $\mathrm{RUN_{HH}}$ shows that the new parameterisation does form NAT PSCs up to about 25 km, like CONTROL and CALIOP observations. However, due to the prescribed number density limits in the NAT scheme, the surface area densities of the NAT PSCs in $\mathrm{RUN_{HH}}$ are lower by an order of magnitude than what is produced in the CONTROL run (resulting in differences of magnitudes up to 4.5 $\mu m^2/cm^3$ in Fig. 5d). Additionally, a Supplementary figure has been included (Fig. S3) to further explain the difference in SAD (see Fig. 2 below). The accompanying text for this revised figure in the revised manuscript now reads:

'To assess the seasonality and magnitude of SAD for the PSCs simulated by RUN_{HH} and CONTROL, Fig. 5 compares the temporal evolution of this quantity for NAT, ice, and liquid droplets (STS and sulphate aerosols) over the extended Antarctic 2008 winter. The smallest contribution to SAD for RUN_{HH} from the three PSC types is from NAT particles, which reach a maximum of around 0.36 $\mu m^2 cm^{-3}$ in mid-June from around 16 to 22 km (Fig. 5a), despite NAT showing the largest spatial coverage (Fig. 3). In contrast, the opposite occurs for liquid droplets and ice in RUN_{HH}, which have much larger SAD (Fig. 5b and c) despite smaller spatial coverage (Fig. 3). Additionally, the peak in SAD for ice particles occurs slightly after that for NAT and at a slightly lower altitude range, while the occurrence of liquid droplets occurs at even lower altitude but with distinct peak. Compared to RUN_{HH}, the SAD for NAT particles simulated by CONTROL is up to an order of magnitude higher (Fig. 5d). Differences in the distribution of SAD values for NAT particles simulated by CONTROL and RUN_{HH} across all grid-cells confirms this (Fig. S3 in the Supplementary Material), showing that CONTROL produces a narrow distribution of SAD for NAT particles concentrated at 1–10 μm^2 cm^{-3}, while RUN_{HH} produces a broader distribution shifted to lower values of 0.01–1 μm^2 cm^{-3}. The

opposite is apparent for ice and liquid droplets (Fig. 5e and f), where CONTROL consistently has lower SAD than RUN_{HH}. Note that results for the other sensitivity simulations were broadly similar to RUN_{HH} (not shown), but with varying SAD magnitudes (see also Fig. S3)'

Thirdly, we thank the reviewer for highlighting the issue on the low HNO_3. We are aware of the issue of low HNO_3 in UKESM and how this likely contributes to the new scheme lacking the early STS formation, and the NAT not reaching the high surface area densities even with the higher number density limits. This issue is already mentioned in the manuscript (Discussion section), which we have now expanded in the revised manuscript to further explain the potential effect of low HNO_3 on STS and NAT formation (new text highlighted in bold):

'The severe underestimation of HNO_3 at the beginning of the PSC season, likely due to a known low-NOy bias in the UKESM1.1 stratosphere (Archibald et al., 2020), means that the improved comparison with observations of gas-phase HNO_3 from MLS using the new scheme does not necessarily imply an improvement in the ability of UKESM1.1 to accurately capture the full extent of HNO_3 depletion and denitrification. This is because UKESM1.1 begins the PSC/winter season with less HNO_3 than observed. ***This bias also affects NAT and STS formation since both are dependent on the partial pressure of HNO_3. The previous thermodynamic equilibrium approach may have partly compensated for this bias in terms of NAT formation by converting all excess gas-phase HNO_3 into NAT, whereas the new scheme slows NAT growth, making the effect of low HNO_3 on NAT formation more apparent. As a result, the reduction in NAT SAD in the new scheme reflects both slower NAT growth and enhanced STS formation, but is likely also amplified by the background HNO3 bias.*** *The origin of this bias requires further investigation'*

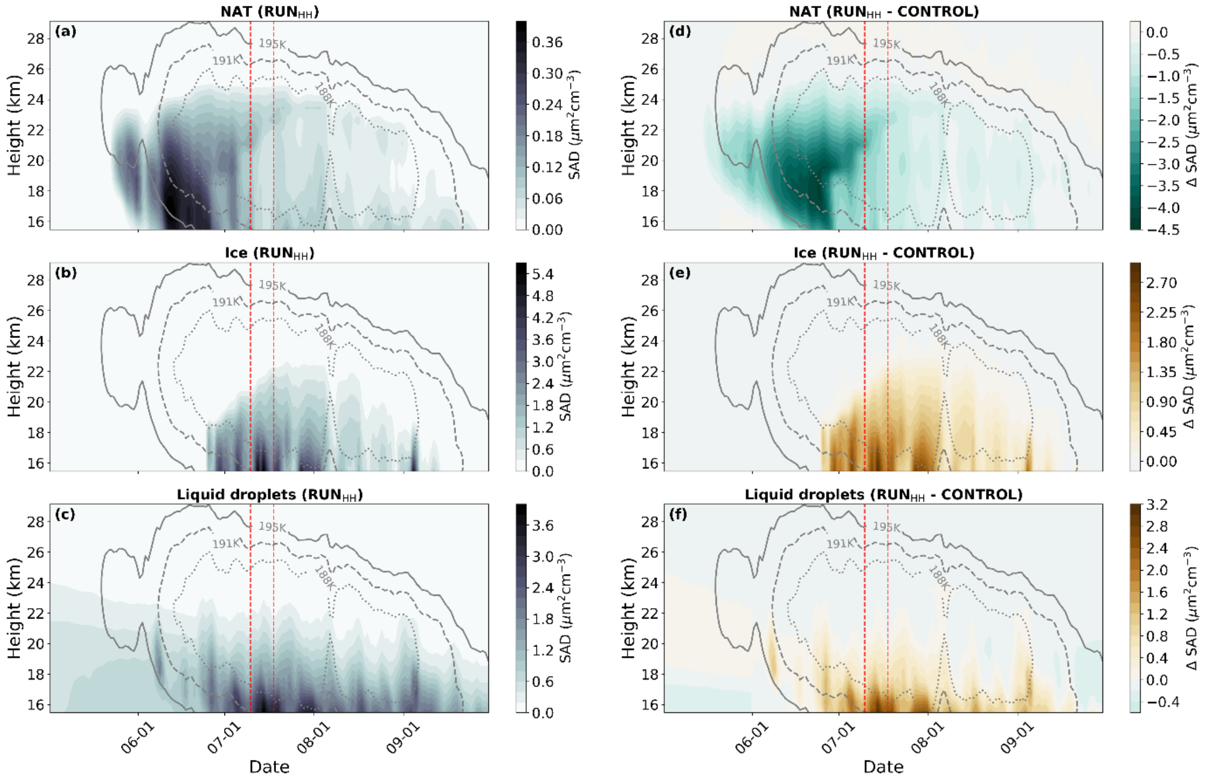


Figure 1: *Time-height plot of zonal mean SAD (shading; $\mu m^2\ cm^{-3}$) for NAT (a), ice (b) and liquid droplets (STS and sulphate aerosols, c) for RUN_{HH} over the Antarctic for the extended 2008 winter season (May to October), averaged across the latitude band 65°S - 90°S. Also shown are the differences between RUN_{HH} and CONTROL for NAT (d), ice (e), and liquid droplets (f). Grey contour lines indicate T_{NAT} (solid), T_{STS} (dashed), and T_{ice} (dotted). The red dashed vertical lines indicate the period July 10–18 2008 examined in Fig. 1. Note that the colour bar for all panels corresponds to different scales.*

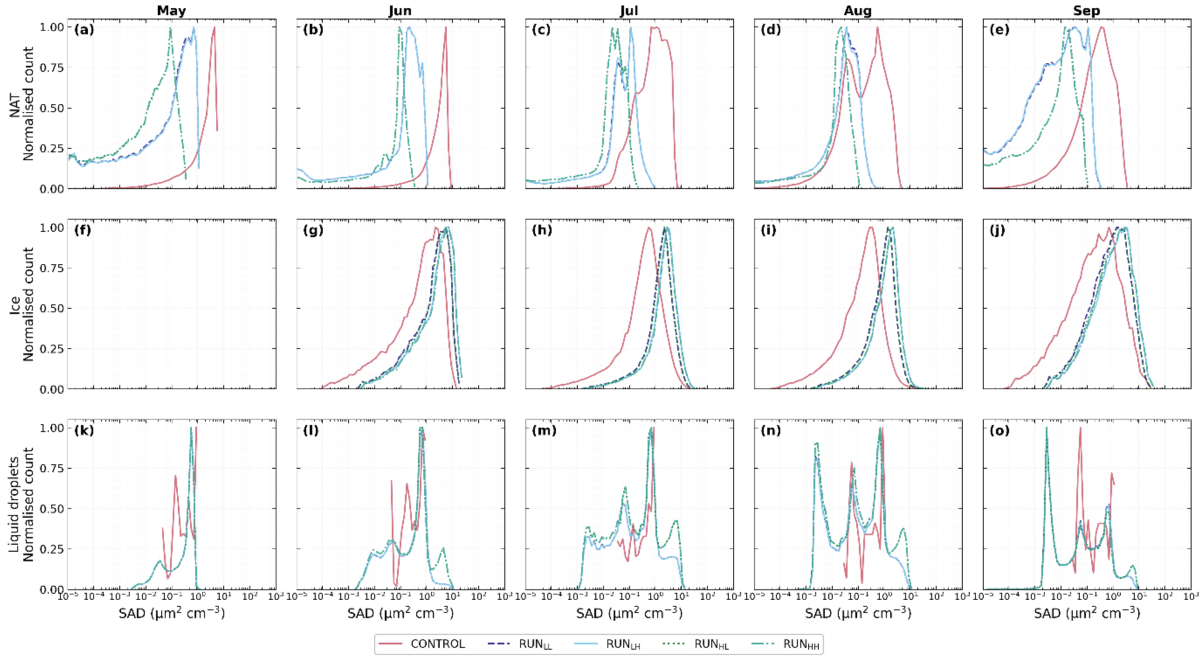


Figure 2: *Distribution of SAD (μm^2 cm^{-3}) for NAT (a-e), ice (f-j) and liquid droplets (STS and sulphate aerosols, k-o) for CONTROL and RUN_{LL}, RUN_{LH}, RUN_{HL}, and RUN_{HH} for the Antarctic for each month of the extended 2008 winter season (May to September), across the latitude band 65°S–90°S.*

2. Section 3 misses dehydration comparison with MLS. Also, what about Arctic denitrification and dehydration?

Reply: We thank the reviewer for this comment and agree that this additional comparison of dehydration and denitrification with MLS would be beneficial in benchmarking the new scheme. These new figures are shown below (Figs. 3, 4 and 5). Figures 3 and 5 are included in the Supplementary Material of the revised manuscript (S5 and S7, respectively) and Fig. 4 is included in Section 3.3 of the revised manuscript (Fig. 8). Note that there is not enough PSC formation for there to be significant observable denitrification or dehydration in the Arctic, which is why we feel that the appropriate place for these plots is the Supplementary Material. The additional analysis shows that:

i) the low HNO_3 bias observable over the Antarctic winter season is also observable over the Arctic winter season, even without the depletion due to PSC formation, and

ii) there is complete overlap in all the model H_2O concentration lines as a result of the new scheme not altering the amount of water that is converted to solid phase to form ice (i.e., this is calculated by the cloud microphysics scheme, which we have not modified). There is a slight underestimation of H_2O in all model runs compared to MLS observations at the start of the Antarctic winter season. Then, less dehydration in the models compared to MLS from July to November results comparable H_2O concentrations to MLS over the 60° S – 70° S and higher H_2O concentrations than MLS over the 70° S – 80° S latitude band.

In addition to the new plots, we have also added the following accompanying text in Section 3.3 :

'Additionally, the same analysis is presented in the Supplementary Material for the extended 2009/2010 Arctic winter at 68 hPa (Fig. S5). Here, there is not the same level of removal of gas-phase

HNO_3, however the same low-bias in the models compared to MLS at the start of the Antarctic winter season is present throughout the entire Arctic winter season.'

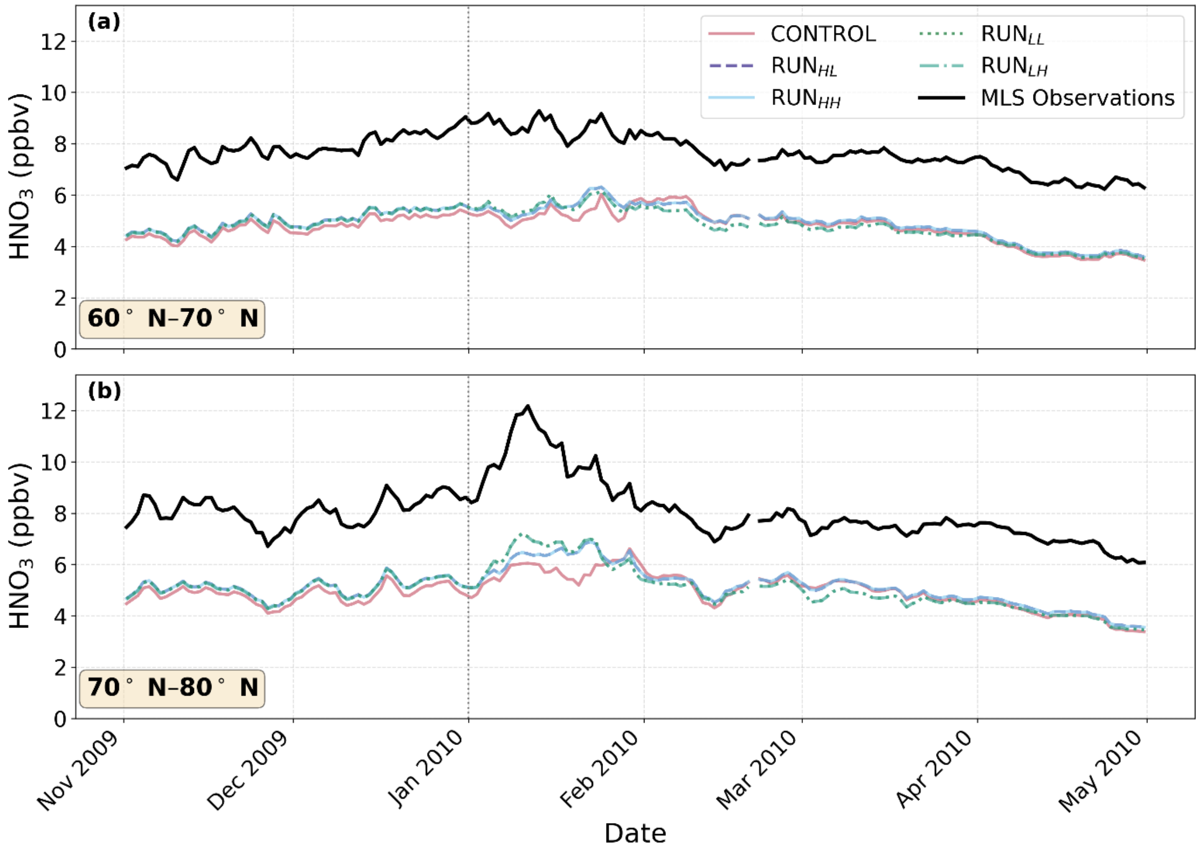


Figure 3: *As Fig. 6 but for the Arctic over the extended 2009/2010 winter season (November to May), averaged across latitude bands 60° N-70° N (a) and 70° N-80° N (b).*

'Stratospheric gas-phase H_2O is additionally analysed in Fig. 8 for the extended 2008 Antarctic winter. All model simulations produce the same gas-phase H_2O concentrations since ice is formed as a part of UKESM1.1's microphysics and is unchanged in the new PSC parameterisation. The new scheme only affects how ice mass mixing ratio is converted to SAD for heterogeneous chemistry. Furthermore, the amount of H_2O that condenses to form STS particles is negligible compared to the gas-phase H_2O concentrations. Compared to MLS, the model slightly underestimates gas-phase H_2O in May and June, but shows weaker dehydration from July to November. As a result, simulated gas-phase H_2O is comparable to MLS over 60° S - 70° S (Fig. 8a) and higher than MLS over the 70° S - 80° S (Fig. 8b). The same analysis is presented in the Supplementary Material for the Arctic winter at 68 hPa (Fig. S7).'

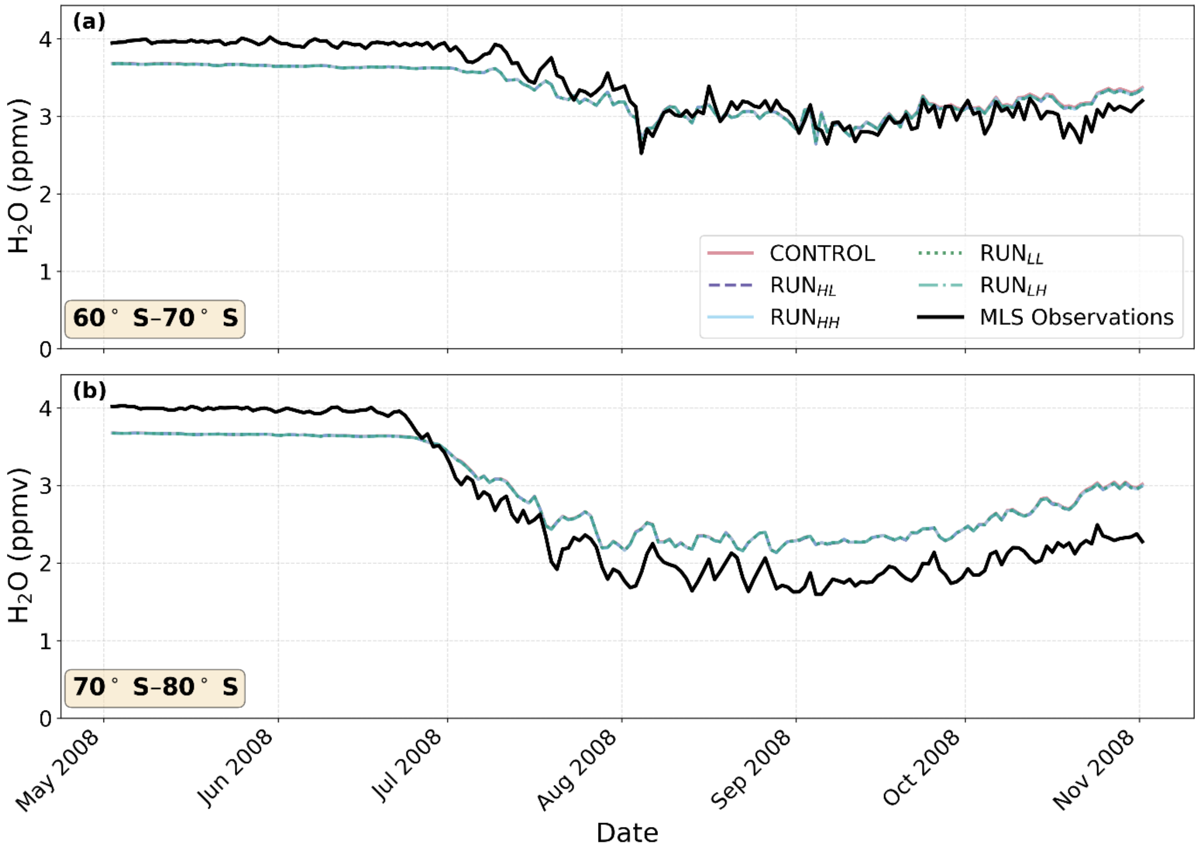


Figure 4: *As Fig. 6, but for zonal mean H_2O (ppmv). Note that there is complete overlap of all model runs (CONTROL and sensitivity simulations).*

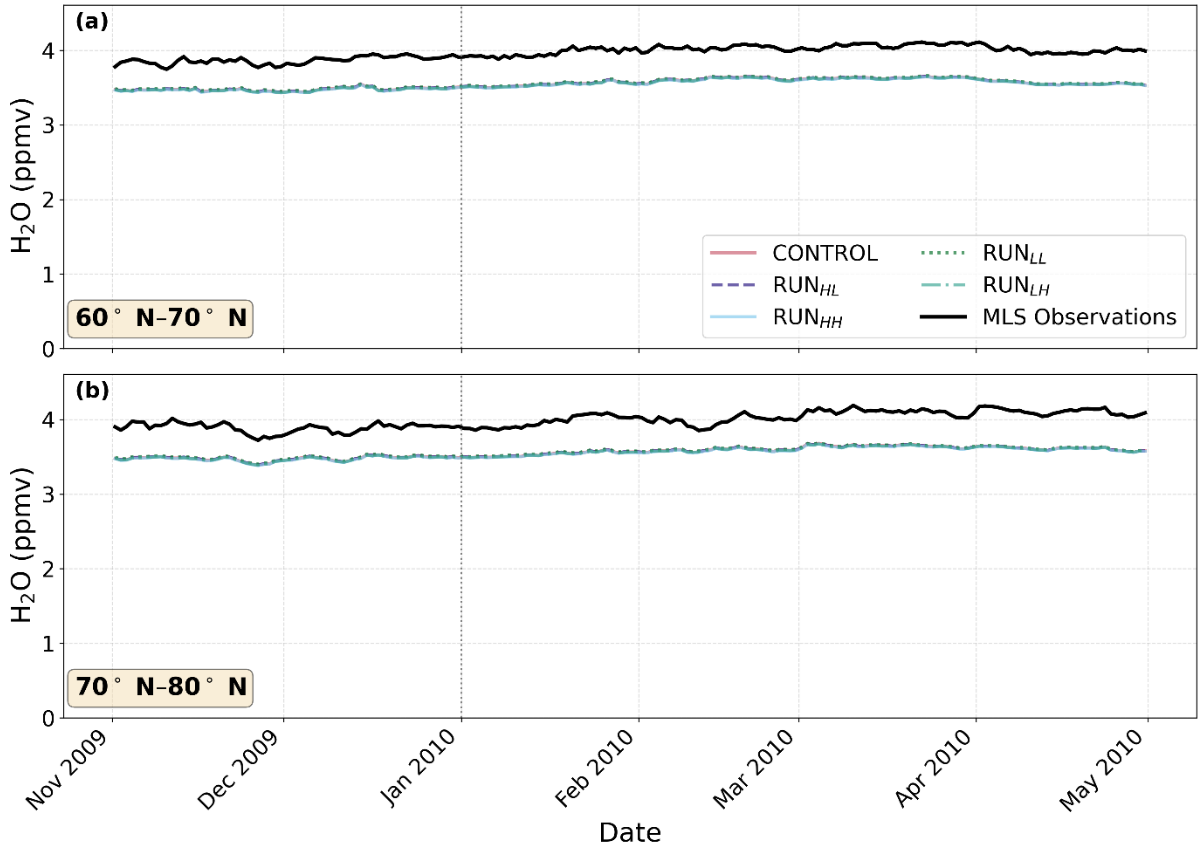


Figure 5: *As Fig. S5, but for zonal mean H_2O (ppmv). Note that there is complete overlap of all model runs (CONTROL and sensitivity simulations).*

3. Figure 1 and Figure 2: The color bars between the model and CALIPSO are completely different. It's hard to compare. You should not use log scale color bar. If your concern is your total counts are different, at least, you should interpolate your color bar proportional to CALISPO color bar. For example, the cyan color is 700 for CALIPSO, you don't have a cyan color but maybe it is like 20 in your color bar. That's a huge difference.

Reply: Thank you for the suggestion, which we agree with. In the revised manuscript, the colour bars for Figs. 1 and 2 have been changed to be linear and to show the total number of counts for a fair visual comparison with the CALIOP optical properties (see Figs. 6 and 7, below). The linear colour bar allows for accurate comparison of the proportion of observations and simulated grid cells within given optical property classifications. We are less concerned with comparing the total number of counts between the model and CALIOP since the model and CALIOP data are on different grids. To clarify this in the revised manuscript, we have included additional text in Section 3.1:

‘***Note that while the values of the counts of observations for CALIOP and grid-cells for the simulations in the 2D histograms cannot be directly compared, the relative abundance can be.’***

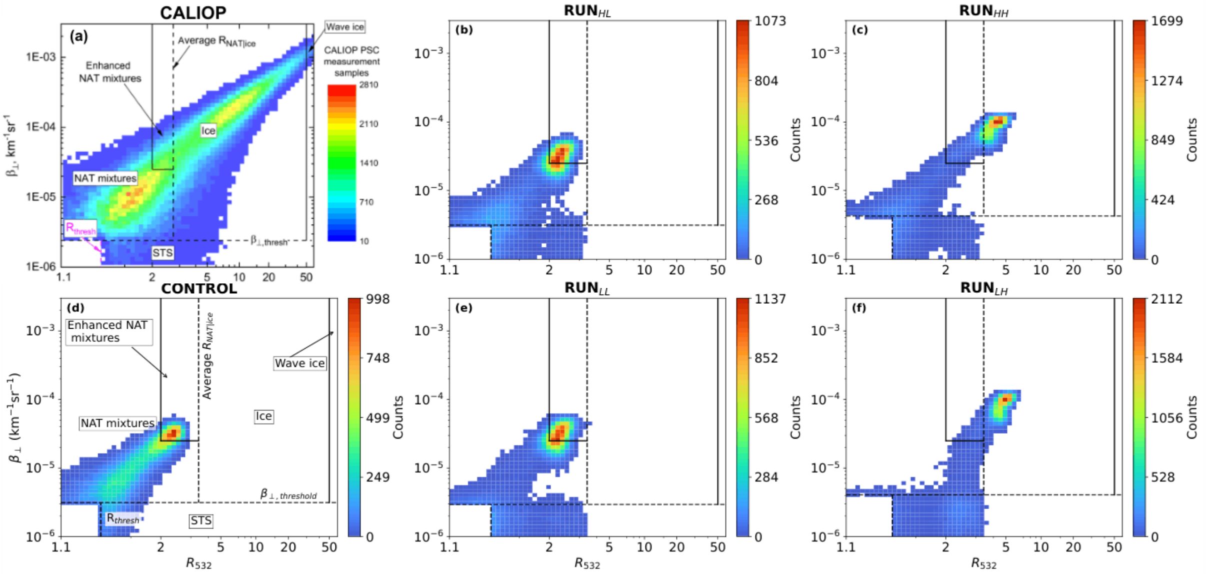


Figure 6: *2D histogram of optical properties from CALIOP (a), RUN$_{HL}$ (b), RUN$_{HH}$ (c), CONTROL (d), RUN$_{LL}$ (e), and RUN$_{LH}$ (f) over the Antarctic for the period July 10–18 2008, latitude range 65°S–75°S, and potential temperature range 475–525 K. The optical properties are plotted in the CALIOP v2 classification coordinate system of R_{532} vs $\beta_{\perp}$, with regions of enhanced NAT mixtures (enhNAT), liquid-NAT mixtures (NATmix), STS, ice, and wave ice indicated. The shading indicates the number of CALIOP observations (a), or the number of model grid points (b–f). Dashed lines indicate the mean $\beta_{\perp}$ and R_{532} thresholds (including measurement uncertainty), as well as the mean $R_{NAT|ice}$ boundary, calculated for CALIOP and each simulation separately. Solid lines indicate the boundaries separating wave ice and ice and enhNAT and NATmix. The regions labelled for CONTROL (d) are the same for all the sensitivity simulations (b, c, e, f).*

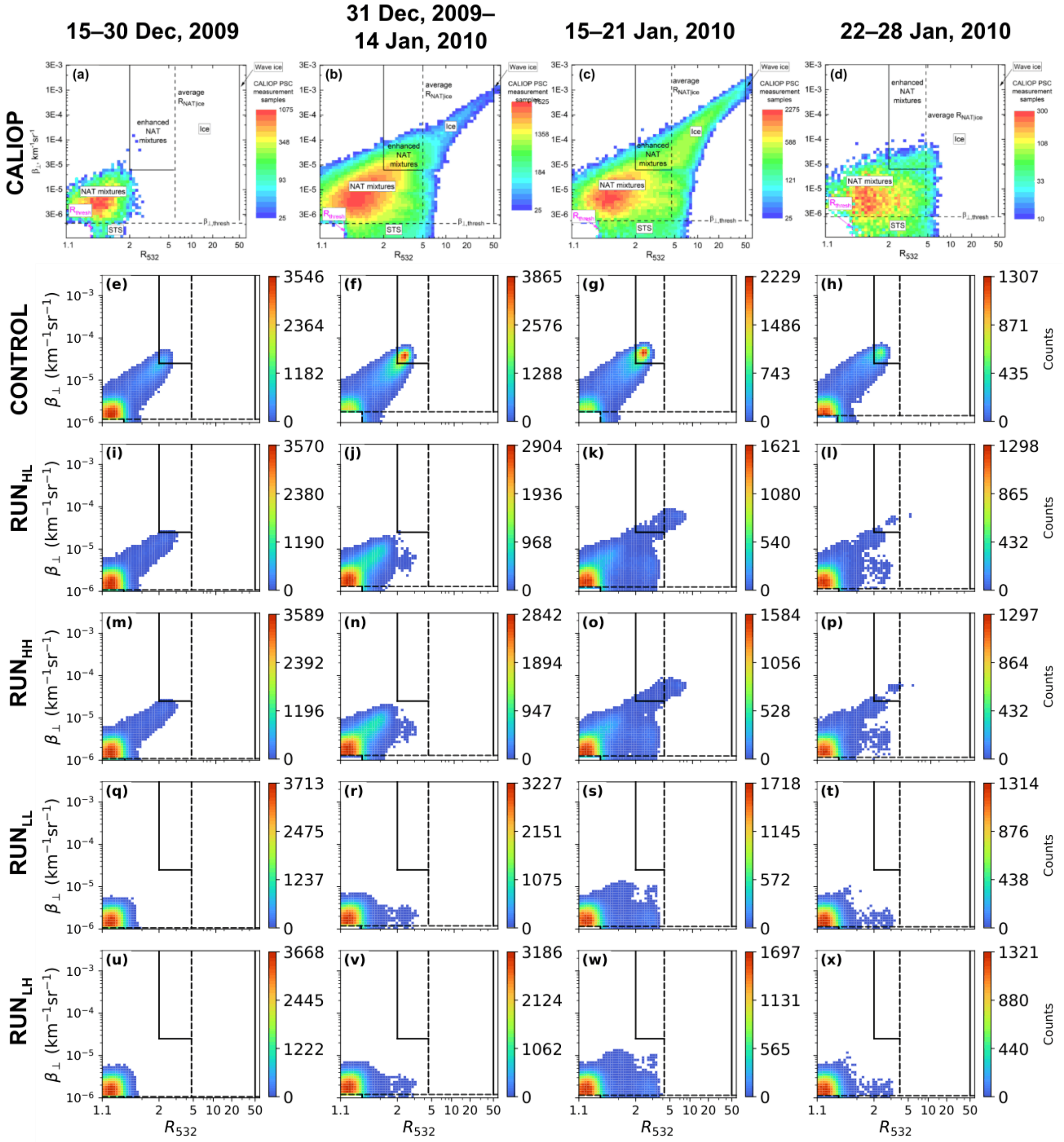


Figure 7: *2D histogram of optical properties from CALIOP (a–d), CONTROL (e–h), RUN$_{HL}$ (i–l), RUN$_{HH}$ (m–p), RUN$_{LL}$ (q–t), and RUN$_{LH}$ (u–x) over the Arctic for the 2009/2010 winter season, covering latitude range 65°N–75°N, and height range of 15–30 km. The results are separated into four distinct periods, comprising 15–30 Dec 2009 (first column), 31 Dec 2009–14 Jan 2010 (second column), 15–21 Jan 2010 (third column), and 22–28 Jan 2010 (fourth column). The results are plotted with the same classification algorithm, shading, boundaries, and regions as Fig. 1*

Other smaller points:

1. Line 26: There's no such a thing called Tsts. STS grows as the temperature decrease, not suddenly at a certain temperature. You could say 191 K is a typical temperature that HNO3 condenses rapidly in a typcial stratospheric chemical conditions (i.e. HNO3 ~ 10 ppb, H2O~ 5ppm and 50 hpa), but HNO3 still condenses slowly on the particle at higher temperatures based on vapor pressure.

Reply: We thank the reviewer for this suggestion and have reworded this sentence to read : '***Under normal stratospheric ambient conditions (i.e., 50 hPa with HNO_3 ~ 10 ppb and H_2O ~ 5 ppm), the typical temperature that HNO_3 rapidly condenses onto background sulphate aerosols is 191 K***

(although HNO_3 still condenses slowly on the background aerosols at higher temperatures if vapour pressures are high enough).'

2. Line 33: Are you talking about H2SO4-H2O sulfate binary? or also other binary?

Reply: We thank the reviewer for the clarification and have updated Line 33 to specify '***liquid binary H_2SO_4-H_2O aerosols***'.

3. Line 76: atmospheric variables -> atmospheric gas variables

Reply: We have updated the wording in this sentence to now read : '***.... providing nearly coincident measurements of atmospheric gas variables ...***'.

4. Figure 1: Zhu et al. 2017a talked about how to adding noise to the model run to have a closer comparison with the CALIPSO data. Adding noise will spread the model results to more categories. You may try that.

Reply: We thank the reviewer for this suggestion. Following Zhu et al. (2017) and Tritscher et al. (2019), background β_{perp} and R_{532} values and observational noise have already been added to the simulated optical properties to bring them closer to the CALIPSO observations. This is described in Sect. 2.3.1 (Eqns. 16–17), where the CALIOP Noise Factor (CNF) is used to calculate and apply noise to both β_{perp} and R_{532} after background values have been added. A random value drawn from a normal distribution with standard deviation $\sigma(\beta_{perp})$ or $\sigma(R_{532})$ is added to each simulated value before classification, consistent with the approach of Zhu et al. (2017).

Referee #2

The paper titled "Assessment of an updated polar stratospheric cloud parameterisation for the UK Earth System Model (UKESM1.1) within the UK Met Office Unified Model (v13.9) using CALIOP and MLS observations" by Isabelle Sangha and co-authors presents a comprehensive and well-executed update of the polar stratospheric cloud (PSC) parameterisation in the UK Earth System Model (UKESM1.1). By incorporating a more physically based representation of PSC formation processes, the authors enhance the model's ability to simulate PSCs. Overall, the manuscript is very well written, logically structured, and carefully researched, with a clear and coherent presentation of methods and results. However, I see some weaknesses in the presentation of the results. Here are my comments:

1. Line 65 ff: PSCs can also be observed with ballon-borne measurements.

Reply: We thank the reviewer for this comment. We have revised this sentence so that it now reads: ***'.. ground-based lidar (Snels et al., 2019) and airborne and balloon-borne in situ observations (Fahey et al., 2001; Voigt et al., 2003)'***

2. Line 130 ff: Peter et al. (1991) is a pretty old publication. Since 1991, much has happened in PSC research and Thomas Peter would not agree anymore to be cited with the statement of "homogeneous NAT nucleation". According to Lowe and MacKenzie (2008), "it appears that, from current knowledge, homogeneous nucleation of NAT/NAD from STS is not significant in the stratosphere, although there is not universal consensus on this subject." Already Koop et al. (1995) showed for bulk solutions that liquid stratospheric aerosol droplets do not freeze under polar winter conditions for temperatures higher than the water ice saturation temperature. Instead, it was discussed in recent years that NAT particles nucleate heterogeneously on foreign nuclei (e.g. Hoyle et al., 2013). For your model simulations, it is not important if NAT nucleates homogeneously or heterogeneously. However, you should discuss this in slightly more detail and cite up-to-date literature here.

Reply: Thank you for suggestion. We feel that it would be out of the scope of this section to discuss in detail the current literature on NAT nucleation as the parameterisation implemented in this study does not account for whether the nucleation is homogeneous or heterogenous. We have however updated the line to read:

'...with a 3 K temperature threshold applied in line with similar NAT parameterisations to account for the supersaturation required for NAT formation (Grooß et al., 2002; Kirner et al., 2011).'

to highlight this threshold is applied in accordance with other models' implementation of this NAT scheme, and not necessarily from the supercooling required for homogeneous NAT nucleation.

3. Figure 1 and 2: It is hard to compare the model results to the observations because of the different color bars.

Reply: Thank you for the suggestion. We have now made the colour bars linear to allow for clearer comparison between model and observations. Please see the response to Referee #1 / Comment #3 (above), which also raised this issue.

4. Figure 3: Why are you not comparing your results to CALIOP surface area densities? This is a product from Pitts et al.

Reply: The SAD product from CALIOP is not recommended to compare with model output as CALIOP-derived SAD are based on published relationships for spherical particles and are therefore only crude estimate for NAT and ice (private communication with Michael Pitts and Lamont Poole). By contrast, the model SAD are calculated from the microphysical properties calculated directly in the model and so comparing with the CALIOP surface area densities would not be a fair comparison. We have included this additional sentence in Section 2.3.1 to explain this choice:

'Comparing directly with the satellite optical properties is preferable to using the CALIOP SAD product derived from these properties, as these estimates are highly uncertain for NAT and ice (Pitts et al., 2018).'

5. Figure 4 and 6: The different colors are difficult to see. They lie on top of each other. I have difficulty distinguishing the shades of blue and green. Could you use dashed and dotted lines, for example?

Reply: Thank you for the suggestion. Both figures have now been updated accordingly so the line styles are now different for the different sensitivity simulations (see Figs. 8 and 9). While remaking the plots we found errors in the previous the MLS latitude band subset: i) MLS only goes from 82° N - 82° S, meaning the 80° S - 90° S latitude band had very few points, and ii) the 60° S - 70° S band had extra points masked due to how the bounds were defined. These errors have been amended in the new figures so 60° S - 70° S includes all the MLS points within this band, and there is no longer an 80° S - 90° S zonal mean panel. This results in the models showing a more significant low bias in HNO$_3$ in the 60° S - 70° S zonal mean than the previous figure indicated which is addressed in Section 3.3 in the revised manuscript:

'While both CONTROL and sensitivity runs have slight gas-phase HNO$_3$ depletion, they consistently underestimate gas-phase HNO$_3$ concentrations, especially in May before significant PSC formation.'

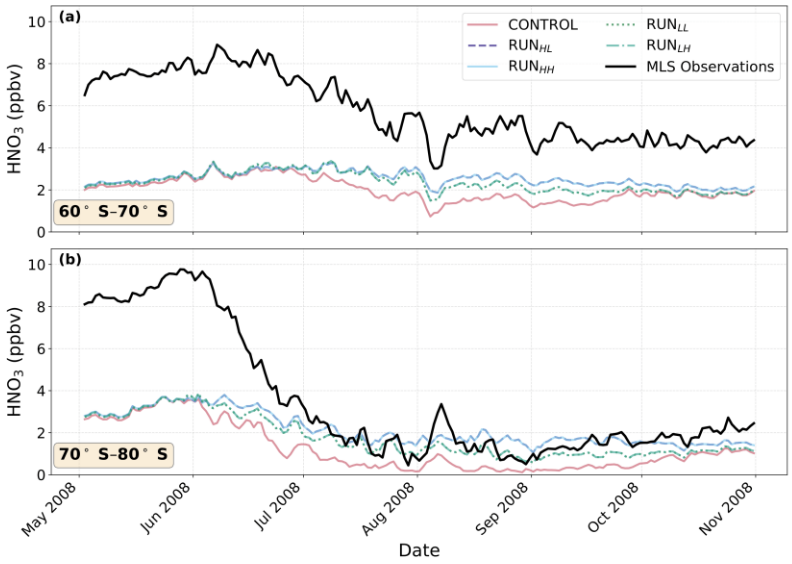


Figure 8: *Zonal mean volume mixing ratio of gas-phase HNO$_3$ (ppbv) at 68 hPa for MLS (black line), CONTROL (pink line) and RUN$_{LL}$, RUN$_{LH}$, RUN$_{HL}$, and RUN$_{HH}$ (purple, blue, green, and teal lines, respectively) over the Antarctic for the extended 2008 Antarctic winter season (May to October), averaged across the latitude bands 60°S–70°S (a), and 70°S–80°S (b). Note that there is almost complete overlap of RUN$_{HH}$ and RUN$_{HL}$, and RUN$_{LL}$ and RUN$_{LH}$, resulting in only RUN$_{HH}$ and RUN$_{LL}$ being visible in the plot.*

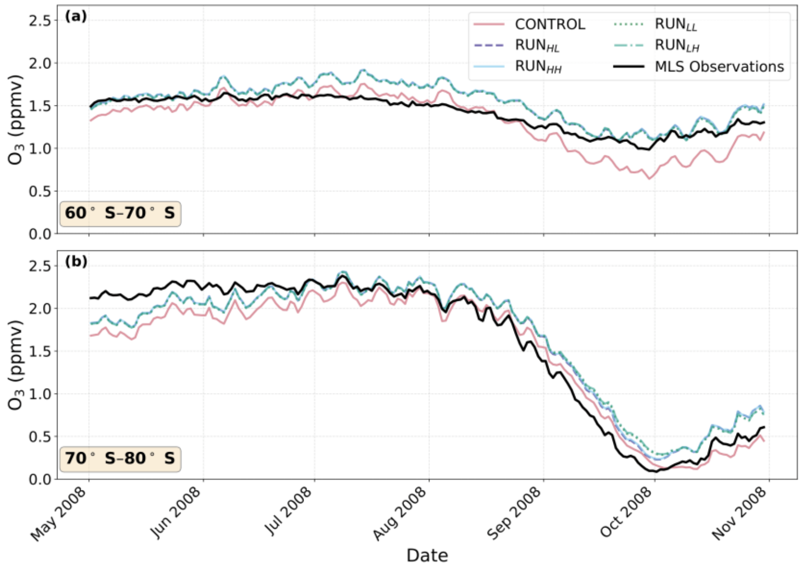


Figure 9: *As Fig. 6, but for zonal mean* O_3 *(ppmv). Note that there is almost complete overlap of* RUN_{HH} *and* RUN_{HL}*, and* RUN_{LL} *and* RUN_{LH}*, resulting in only* RUN_{HH} *and* RUN_{LL} *being visible in the plot*

6. Generally, I have the impression that the comparison with CALIOP could be much more detailed.

Reply: We thank the reviewer for this insight, and while we decided it was best to not compare the SAD of PSCs from CALIOP and the model simulations directly (see Referee #2 / Comment #4, above), we have included additional model-only analysis for the revised manuscript for RUN_{HH} and CONTROL on the occurrence frequency of each PSC type over Antarctica for each month from May – September averaged over the 5-year period from 2007 – 2011 in Section 3.2, which we think provides more details (see Figs. 10 and 11). In the revised manuscript, this new analysis is Figs. 3 and 4. Additionally, we have included a new figure of the areal coverage of each PSC type in RUN_{HH} and CONTROL over the 2008 Antarctic winter in the revised manuscript's Supplementary Material materials (see Fig. 12; Fig. S2 in revised manuscript), which is also mentioned in Section 3.2.

The new text in section 3.2 of the revised manuscript that explains these additional figures is as follows:

'Figures 3 and 4 show the average PSC occurrence frequency over Antarctica at 20 km for NAT, ice, and STS for every month of the extended winter season over the entire 5-year simulation period of 2007–2011 for RUN_{HH} ***and CONTROL, respectively.*** RUN_{HH} ***was chosen as it shows the largest spread of optical properties compared with CALIOP (Fig. 1). Here,*** RUN_{HH} ***has high NAT occurrence frequency across Antarctica from June to August (values reaching 1), with lower frequencies in May and September (reaching 0.2). High occurrence frequency of ice (reaching 0.9) also occur during July and August, and are mainly concentrated over the centre of Antarctica. In contrast, STS only occurs from June to August in*** RUN_{HH}***, particularly over Western Antarctica and the Antarctic Peninsula (i.e., near the edge of the stratospheric polar vortex), with much lower occurrence frequencies (reaching 0.2) - although this STS pattern is consistent with CALIOP observations (Pitts et al., 2018, Fig. 19 therein). CONTROL shows broadly similar patterns in both NAT and ice occurrence frequency as*** RUN_{HH}***. The main differences are higher NAT frequency in May and September, reaching 0.7, and slightly lower ice frequencies in July and August. As CONTROL does not form STS, these results are not included in our comparison. The other sensitivity simulations show similar behaviour to*** RUN_{HH} ***(not shown). These results are further supported by the seasonal evolution of Antarctic PSC areal coverage for the extended 2008 winter season (see Fig. S2 in the Supplementary Material). Both CONTROL and*** RUN_{HH} ***show broadly similar areal coverage for NAT and ice PSCs, although there are slight differences in timing and vertical distribution. STS only forms in*** RUN_{HH} ***and starts having notable areal coverage later in the season and at a lower altitude***

than NAT. Both simulations show broadly similar total PSC areal coverage over the entire winter season, although CONTROL has larger extent at higher altitude and RUN$_{HH}$ at lower altitude.'

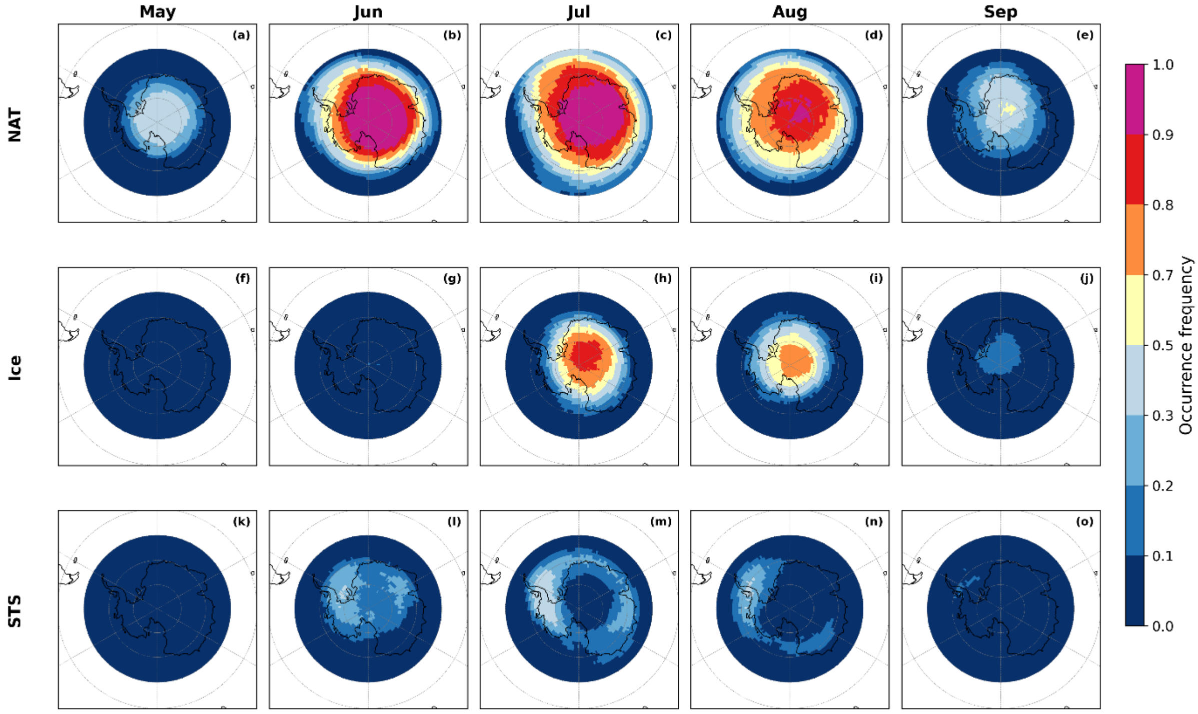


Figure 10: *Monthly average PSC occurrence frequency over Antarctica of NAT (a-e), ice (f-j) and STS (k-o) for May-September (left to right) over the 5-year period from 2007 to 2011 at an altitude of 20 km for RUN$_{HH}$. Occurrence frequency is defined as the number of days with PSC occurrence over the 5-year period for each month, divided by the total number of days over the 5-year period for each month.*

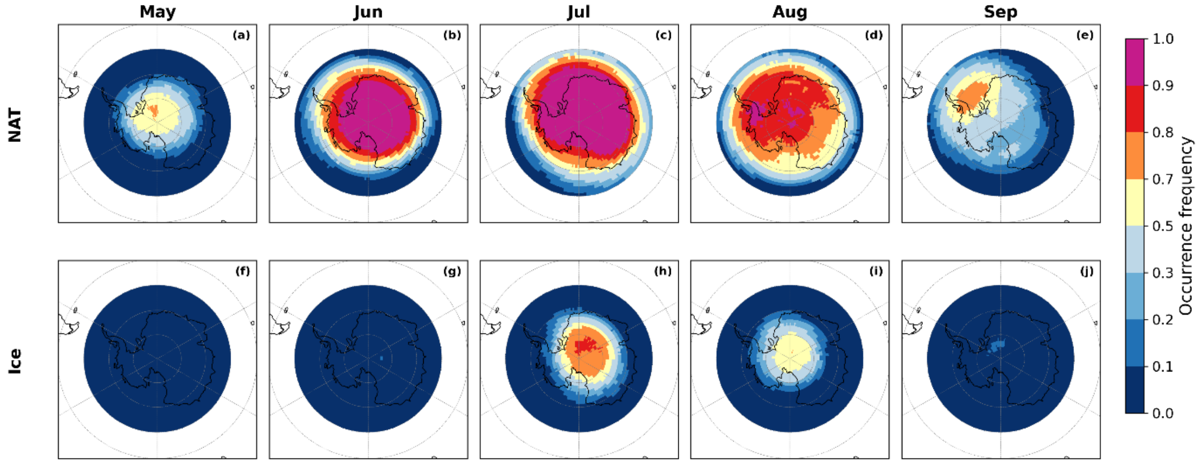


Figure 11: *As in Fig. 3, but for CONTROL. Note that as CONTROL does not form STS, only results for NAT and ice PSCs are shown.*

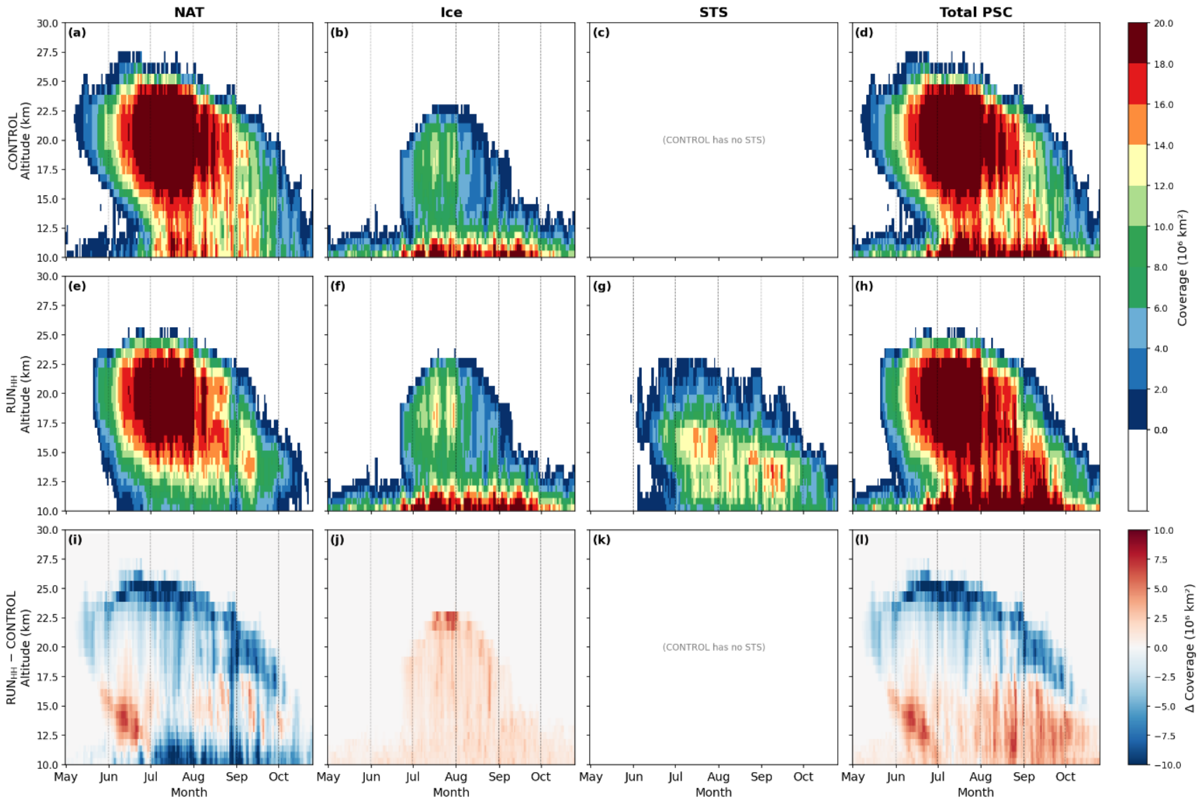


Figure 12: *Time-height seasonal evolution of the PSC areal coverage (km^2) for NAT (a,e), ice (b,f), STS (g), and total (d,h) PSCs for CONTROL (a,b,d) and RUN_{HH} (e-h) for the Antarctic over the extended 2008 winter season (May to October) and averaged across the latitude band 60° S–90° S. Also shown are the differences in PSC areal coverage between RUN_{HH} and CONTROL for NAT (i), ice (j) and total (l). Note that since CONTROL does not simulate STS, panels c and k are left blank.*

7. Line 329 ff: The model's underestimation of HNO3 is extreme. Should this not be solved first? It could bias the entire analysis.

Reply: We agree with the reviewer that the underestimation of HNO_3 in the model is considerable and likely affects the PSC formation. This issue was also raised by Referee #1 / Comment # 1. Sorting this issue out is way beyond the scope of this investigation. However, the Summary section of the manuscript clearly outlines this issue, and remarks that solving it should be a priority for the UKESM modelling community.

8. What does H2O look like? There are no plots referring to H2O!

Reply: Following the comment from Referee #1 / Comment #2, we have added additional analysis to the revised manuscript that compares model v MLS-based H2O. Please see the reply to Referee #1 / Comment #2 and Figs. 4 and 5 (above).

9. Line 406: “high-number-ddensity”

Reply: Thank you. We have fixed the spelling error to now read ‘***high-number-density***’.

10. In summary, I have the impression that the evaluation against CALIOP satellite observations and Aura MLS measurements is not yet fully convincing. As it stands, the plots tend to give the impression that the agreement between model and observations is weaker than intended. The plots should be improved to strengthen the overall conclusions.

Reply: In the revised manuscript we have now included additional analysis, including i) in section 3.2 the occurrence frequency of each PSC type (in response to Referee #2 / comment #6), ii) in section 3.3 comparison of denitrification in the Arctic and dehydration in both the Antarctic and Arctic against MLS observations (in response to Referee #2 / comment #8 and Referee #1/Comment #2). The inclusion of these additional plots expands the evaluation of the new scheme and strengthens the overall conclusions / purpose of the manuscript, which is to accurately benchmark the new scheme and document its behaviour.

Referee #3

This paper compares new polar stratospheric cloud (psc) models to the standard model (control) used in the UK Earth system model. Assessments of improvements to the representation of pscs and impacts on nitric acid and ozone are made through comparisons with CALIOP and MLS observations. The paper is mostly ready for publication, but there are a few serious issues that need to be addressed before that.

Here by line, table, figure number are comments the authors should consider as they prepare a revision.

1. 30-35 To what extent is the difference between NAT and STS/sulfate aerosol activation chemistry due to the chemistry, or the surface area available? Generally STS/sulfate aerosol are significantly smaller and more numerous than NAT thereby usually providing significantly more surface area.

Reply: Thank you for this comment. Both available surface area density and reactivity per unit surface area result in STS/sulphate aerosols activation chemistry being more effective than NAT activation chemistry. To clarify what we are referring to we have changed the wording in the opening paragraph of the revised manuscript to be:

'… liquid droplets (e.g., STS) are comparable to, or more effective than, solid NAT and ice particles in activating chlorine based on reactivity per unit surface area (Ravishankara and Hanson, 1996).'

2. 110 The sedimentation velocity is fixed for all pressure levels? This is clearly wrong but that's what the text indicates. Some explanation should be added.

Reply: We thank the reviewer for this comment. The previous sedimentation scheme had a fixed sedimentation velocity at all model levels as described in Line 110 and Morgenstern et al. (2009). However, the new scheme interactively calculates the sedimentation velocity and we have adjusted the text in Section 2.1.2 to read: '***…NAT sedimentation is determined interactively, with the fall velocity calculated for each bin based on Stokes' law…***' to highlight the sedimentation velocity is calculated based on Stokes' law, which is a function of pressure.

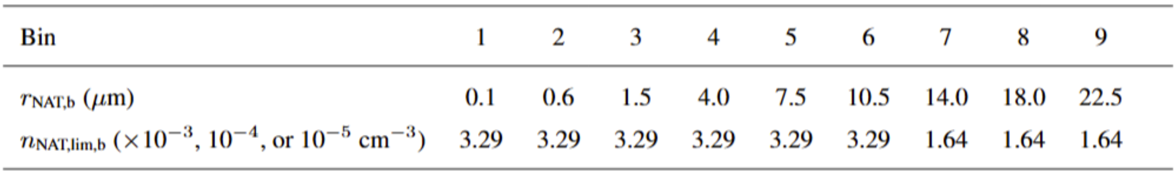
3. 155 Isn't this the effective radius?

Reply: We thank the reviewer for clarifying this. The radius used in the optical calculations is the effective radius, not mean radius. We have updated the relevant text in section 2.1.2 to read '***The effective radius of the STS particles …***'

4. Table 2 is confusing. A is set to orders of magnitude. But then the maximum number densities are given to three significant figures: once in the table description, and then a different set of three significant figure number densities, times some value of A, in the table. In the table the n_NAT,lim,b exceed the total number density for bins 1-6. The point of this table is not at all clear, and is certainly not going to be understood by any reader outside of the authors' modeling group.

Reply: We thank the reviewer for highlighting this and have updated Table 2 in the revised manuscript to more clearly show the different magnitudes of the number density limits of the size bins across the sensitivity simulations and have updated the caption to more clearly explain this (see Table 1 below). We have additionally updated the text in Section 2.3.2 to better explain the number density limits of the bins in the sensitivity simulations:

'In the implementation of the same kinetic NAT parameterisation used here, Kirner et al. (2011) and Weimer (2019) use $n_{NAT,lim,b}$ values of the order of 10^{-5} cm^{-3} based on measurements by Fahey et al. (2001). Here, we additionally investigate the effect of increasing the $n_{NAT,lim,b}$ values to the order of 10^{-4} and 10^{-3} cm^{-3} to cover the range of number densities observed (Fahey et al., 2001; Biele et al., 2001) and used to constrain CALIOP observations (Pitts et al., 2011). The size bins and different orders of magnitudes used for the $n_{NAT,lim,b}$ are summarized in Table 2 and result in values of the total maximum number density limit, $n_{NAT,max} = \sum_b n_{NAT,lim,b}$, of 2.47×10^{-2} cm^{-3}, 2.47×10^{-3} cm^{-3}, and 2.47×10^{-4} cm^{-3} (for $n_{NAT,lim,b}$ values of the order of 10^{-3}, 10^{-4}, and 10^{-5} cm^{-3}, respectively).'

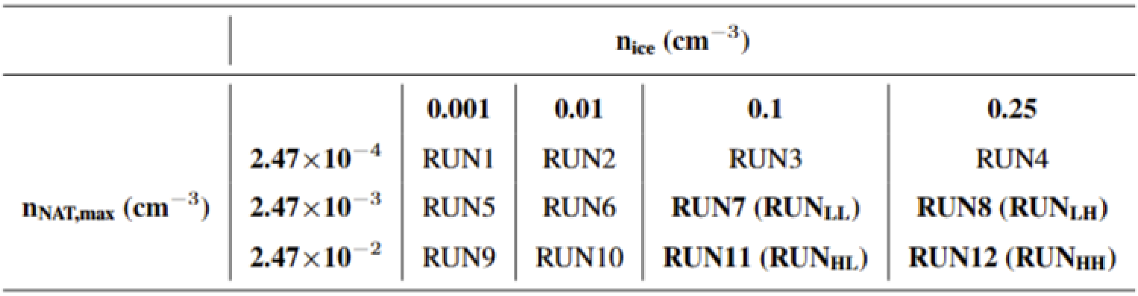


Bin	1	2	3	4	5	6	7	8	9
$r_{\mathrm{NAT,b}}$ (μm)	0.1	0.6	1.5	4.0	7.5	10.5	14.0	18.0	22.5
$n_{\mathrm{NAT,lim,b}}$ ($\times 10^{-3}$, 10^{-4}, or 10^{-5} cm^{-3})	3.29	3.29	3.29	3.29	3.29	3.29	1.64	1.64	1.64

Table 1 *Average radius ($r_{NAT,b}$) and number density limit ($n_{NAT,lim,b}$) of NAT particles in each size bin (b) The order of magnitude of the $n_{NAT,lim,b}$ values depends on the sensitivity simulations and is set to either 10^{-3}, 10^{-4}, or 10^{-5}.*

5. Table 3. Now in the simulations, the number densities are 2 significant figures which still seems a bit exact, but okay, and they all correspond to the numbers in the title of Table 2, rather than the numbers in Table 2.

Reply: For consistency between Table 2 and Table 3, we have included these to 3 significant figures and adjusted the caption of Table 3 to make it clear (see Table 2, below). With the response to Comment #4 above, we hope it is clear that the numbers in Table 3 are the total maximum number density limit, $n_{NAT,max} = \sum_b n_{NAT,lim,b}$, for each simulation as described in Section 2.1.2.



		n_{ice} (cm^{-3})			
		0.001	**0.01**	**0.1**	**0.25**
$n_{NAT,max}$ (cm^{-3})	2.47×10^{-4}	RUN1	RUN2	RUN3	RUN4
	2.47×10^{-3}	RUN5	RUN6	**RUN7 (RUN$_{LL}$)**	**RUN8 (RUN$_{LH}$)**
	2.47×10^{-2}	RUN9	RUN10	**RUN11 (RUN$_{HL}$)**	**RUN12 (RUN$_{HH}$)**

Table 2 *Parameter settings for total maximum NAT number density limit ($n_{NAT,max}$) and ice number density (n_{ice}) in cm^{-3} for each of the sensitivity simulations. RUN7, RUN8, RUN11, and RUN12 are selected for analysis in this study and are highlighted in bold and renamed according to whether $n_{NAT,max}$ or n_{ice} is relatively high (H) or low (L) with respect to the four chosen simulations.*

6. Figure 1 caption, line 4 should be … indicate the beta_perpendicular and R_532 thresholds.

Reply: The caption now reads: ***'..dashed lines indicate the mean $\beta_\perp$ an R_{532} thresholds…'*** (Please refer response to Comment #8 below).

7. Figures 1 and 2 have serious problems. For ease of viewing the panel size of the model results, control and all the runs should match the CALIOP observation panels. This can be done by reducing the size of the model figures or moving the color bar explanation to the top or bottom of the CALIOP figures and increase their size to match the model results.

Reply: Thanks for the suggestion. The panel sizes of the model results have now been reduced so that they are closer to size of the CALIOP observation panels. Additionally, the colour bars have been included for each model results panel. Please see Figs. 6 and 7 (above), and also the response to Referee #1 / Comment #3.

8. The more serious problem is that the boundaries established by Pitts et al. are not reproduced in the model figures. In the Pitts et al. Antarctic observations, minimums for ice (beta_perp, R_532) are: (2.3e-6, 2.8). In the model panels these minimums are (3e-6, 3.4). Further the NAT mixture boundary from Pitts is beta_perp = ~2.4e-5, while in the model panels this line is at 2e-5. In addition the color scales for the number of counts do not match. In Pitts et al the color scale (red/orange) peaks at 2810. In the model panels the same color scale is at 1000, which for the Pitts et al. panel is in the transition from cyan to green. These differences make it difficult to compare the model panels with the observation panel, yet isn't that the point of the figure? If it is then a better job needs to be done to reproduce the boundaries and color scales in the model panels so they match, to the extent possible, the observations. In model, and log, space these differences are significant and confounds an easy comparison of the figures. These

discrepancies are somewhat lessened in the Arctic cases, but still to the extent possible efforts should be made to make them match.

Reply: We thank the reviewer for this careful comparison. They are correct that the boundaries in the model panels did not match those of Pitts et al. (2018), and this comment revealed an error in our figures: Figures 1 and 2 had been plotted using the same mean β_{perp} and R_{532} thresholds for all model histogram panels (calculated as the mean over all Antarctic simulations) rather than the simulation-specific mean for each panel. This has been corrected, and each panel now shows the mean threshold calculated only from the grid cells contributing to that panel. We have additionally improved the caption to more clearly describe that the thresholds are calculated for each simulation, so Figure 1's caption in the manuscript now reads: ***'…Dashed lines indicate the mean $\beta_{\perp}$ and R_{532} thresholds (including measurement uncertainty), as well as the mean $R_{NAT|ice}$ boundary, calculated for CALIOP and each simulation separately….'*** The revised figures (and captions) are shown above as Figs. 6 and 7. See also response to Referee #1/Comment #3.

We note that some difference between the model and CALIOP threshold lines remains expected. As described in Sect. 2.2 (Eqns. 14–15) and Sect. 2.3.1 (Eqns. 16–17), both the CALIOP and model thresholds are calculated per observation or grid cell respectively, and the mean values plotted in each panel will therefore naturally differ. We have clarified this in the manuscript by adding:

"As with the CALIOP observations, each model grid-cell optical properties must exceed the $\beta_{\perp,threshold}$ or $R_{532,threshold}$ based on Eqns.14 and 15, respectively, with σ calculated for each grid-cell using Eqns.16 and 17."

The same applies to the $R_{NAT|ice}$ boundary, which is calculated dynamically for CALIOP using MLS data (Sect. 2.2, lines 199 - 200) and interpolated to the UM grid.

Regarding the colour scale, this has been changed from logarithmic to linear, and the count values are now labelled explicitly on each panel to allow for direct comparison (Response to Referee #1 / Comment #3 and Referee #2 / Comment #3).

9. The discussion of these figures devotes just once sentence to the striking result that none of the model scenarios reproduce the large beta_perpendicular and R_532 space covered by CALIOP in the ice region. In contrast the model results occupy a very small fraction of this space in a very narrow beta_perp and R_532 space. In the Antarctic this space is only penetrated with n_ice high while in the Arctic only if n_nat is high.

Reply: Thank you for pointing this out. We have included additional explanation in Section 3.1 to the address the inability to reproduce the larger optical property values:

"The sensitivity simulations with higher n_{ice} (RUN_{HH} and RUN_{LH}) have optical properties extending into the ice region, however they don't cover the same range of optical properties in the ice region as the CALIOP observations. While CALIOP has a substantial number of observations with R_{532} values ranging from 5 to 20, the maximum R_{532} value achieved by the sensitivity simulations is around 6. Additionally, none of the sensitivity simulations have points with high enough R_{532} values to be classified as wave ice"

Additionally, we include more discussion and likely causes of this missing region of the optical space in the Summary section, addressed Referee #3/ Comment #10.

10. 308-309 on what basis is the parenthetical clause justified?

Reply: The parenthetical clause is no longer in the revised manuscript due to the new analysis added. However, this comment is addressed with the additional text added at the beginning of Section 3.2:

'RUN_{HH} was chosen as it shows the largest spread of optical properties compared with CALIOP (Fig. 1).'

11. 307-310, Figure 3. Now we find that in comparison of surface area to the control all the new model results are similar in surface area, independent of n_nat and n_ice? This figure also does not provide any assessment of the improved psc model. It is just a comparison of one model run with a control. Presumably there are no corresponding surface area densities available from CALIOP? If there were that would be a much more useful comparison.

Reply: Thank you for the suggestion. We decided to not include a comparison with CALIOP SAD product (see Referee #2 / Comment #4), however we have included additional analysis comparing the occurrence frequency and PSC areal coverage (see Referee #2 / Comment #6) in Section 3.2.

We agree that it is not clear why the surface area analysis is not presented for all the sensitivity simulations and have added text to Section 3.2 of the manuscript and an additional Supplementary figure (S3) to clarify this (see Referee #1/ Comment #1, Fig. 2 above):

“Note that results for the other sensitivity simulations were broadly similar to RUN$_{HH}$ (not shown), but with varying SAD magnitudes (see also Fig. S3).”

12. Figure 4 The colors for the 4 runs need to be modified so they can be differentiated. Now they are just some different shade of blue, so the reader can only easily see the MLS observations and the control run. The rest are a guessing game and the legend is no help. Better more differentiated colors are an option, so are different lime styles.

Reply: We have now changed the line styles so that the four sensitivity simulations have different line styles to make them easier to differentiate (see figure in Referee #2/ Comment #5).

12. Figure 5 For a real comparison of run HH to control wouldn’t both the solid and liquid pscs have to be added? It would be easy and useful to include such a line.

Reply: Thank you for highlighting this. We have made the figure caption for Fig. 5 (now Fig. 7 in the revised manuscript) clearer to explain it is a stacked area plot where the sum of green and blue is the total solid and liquid phase HNO_3. The new figure caption now reads:

“Zonal mean volume mixing ratio of solid- and liquid-phase HNO_3 (ppbv) at 68 hPa for CONTROL and RUN$_{HH}$ over the Antarctic for the extended 2008 winter season (May to October), averaged across latitude bands 60°S–70°S (a), and 70°S–80°S (b). Solid-phase HNO_3 for CONTROL is shown with a pink line. The solid- and liquid-phase HNO_3 for RUN$_{HH}$ are shown as a stacked area, with solid-phase in green and liquid-phase in blue. Note that liquid-phase HNO_3 is not shown for CONTROL, as this is not simulated by the original PSC parameterisation. Also, the panels have different scales”

13. 341 It is a significant stretch to suggest that run HH starts partitioning hno3 to the solid phase prior to the control. Rather the timing of onset of both, and the amounts, are far more comparable than they are different, until the control takes off. To suggest differently is to use information not accessible in Fig. 5.

Reply: We thank the reviewer and agree that RUN$_{HH}$ does not start meaningfully partitioning HNO_3 to solid phase before CONTROL. The line has now been updated. It reads:

“The results show that RUN$_{HH}$ starts partitioning HNO_3 from the gas-phase to solid-phase at roughly the same time as CONTROL (early June), but that the amounts of solid-phase HNO_3 quickly surpasses RUN$_{HH}$.”

14. Figure 6 Same criticism as Figure 5 concerning the choice of colors to represent the model runs. There are a lot of options than just different shades of blue, although in this figure there is almost no discrimination between the two model runs shown, _HH and _LL.

Reply: We have changed the line styles so that the four sensitivity simulations have different line styles (see Referee #2/ Comment #5).

15. 368 Figures 1 and 2 do not show surface area densities.

Reply: We have changed the wording to make it clear that we are not saying that Figure 1 and 2 are showing surface area densities and are instead referring to change within the model code:

"Comparisons with CALIOP observations for the 2008 Antarctic and 2009/2010 Arctic winters demonstrates that the new parameterisation scheme achieves closer agreement with CALIOP observations by increasing the number of points classifying as STS and ice and capturing seasonal variability in PSC types (Figs. 1 and 2)."

16. 371 Do the authors mean 70-80 S? In the 60-70 S band, calling the slightly higher hno3 in the model runs, compared to the control, an improvement with MLS hno3 is a stretch.

Reply: Thanks for this comment. This line no longer exists in the revised manuscript, it instead reads:

"The updated scheme leads to modest improvements in gas-phase HNO_3 compared to MLS, particularly within the polar vortex, although all model simulations significantly underestimate HNO_3 (Fig. 6). "

17. 391-414 The conclusions conveniently skip over the most glaring discrepancy between the model and CALIOP: The region tagged as ice bounded by beta_perp 2.3e-6 to 2e-3 and R_532 from 2.8 – 50. The region only appears in Antarctic runs if n_ice is high, but even then the space occupied is limited to bet_perp 4e-5 to 2e-4 and R_532 3.3 to 5.3, and only appears in the Arctic if n_nat is high. This discrepancy rates just half a sentence in the conclusions. While the rest of that paragraph, and two more (!), are devoted to wave ice, a much smaller region of the beta_perp, R_532 sample space, with only a smattering of CALIOP observations. It is not a surprise that an Earth systems model cannot reproduce wave ice. What is a surprise is that the authors choose to discuss at length such a small region of psc space. The much more important difference is the difficulty in the models producing much regular ice, a much larger region of the sample space, with numerous CALIOP observations, yet very little penetration by the new psc models. The current emphasis on discussing wave ice is entirely out of place.

Reply: Thank you for this suggestion. We agree with the reviewer that the conclusion has too much emphasis on the lack of wave ice and not enough on the lack of ice. We have rewritten the conclusions to focus more on the missing ice optical properties in the model simulations and reduce the section on wave ice. Specifically:

"The CONTROL and sensitivity runs produced less ice than observed by CALIOP and failed to form wave ice (Figs. 1 and 2). This likely arises from the prescribed ice number density, which constrains the ice particles to a narrow range of SAD (Fig. S3). This may limit the model's ability to reproduce the broader range of optical properties observed by CALIOP."