Reviewer 1

[C-1] Review of Goto et al, 2023

This paper compares results from a single moment cloud microphysics scheme (NSW6 simulations) with those from a double moment scheme (NSW6) in global 6-year 14km resolution simulations. The models are also evaluated using observations and other models.

Both versions perform well against observations of LWP and precipitation in a general sense although a deeper analysis of this evaluation would have been useful – e.g., focussing on certain important areas such as stratocumulus regions that are important for the aerosol-cloud interaction radiative forcing.

It is claimed that NDW6 performs better than NSW6, but I don't think that is justified. It is clear that NSW6 performs better in the tropics and NDW6 at high latitudes. It would be better to be upfront about this and to discuss/examine some of the reasons why this is the case.

Hence, the justification for choosing NDW6 over NSW6 is not that strong from the aerosol results. Perhaps this would be better justified by the likely better representation of cloud microphysical processes in NDW6 (e.g, a better representation of fall speeds, hydrometeor separation, droplet evaporation, etc.?). And the large difference in ERFaci are also a good reason to consider NDW6 - although which one is more accurate is unknown... It might be worth mentioning and considering that the LWP response to aerosol could be constrained/evaluated to some degree with a short simulation of a well-observed volcanic eruption, e.g., like the Holuhraun eruption (Malavelle, Nature, 2017; doi:10.1038/nature22974)

There is a large difference in the aerosol-cloud interaction forcing between the NSW6 and NDW6 simulations, but there is less analysis and discussion devoted to that in the paper. For example, there is no analysis about potential differences in the Twomey effect and cloud lifetime adjustments. E.g., how big might these be in NDW6 vs NSW6? Can you compare the PI to PD CDNC differences for NDW6 to the CCN differences in NSW6? Can you compare changes in the cloud fraction (particularly low altitude clouds) – (this should be possible if the simulations were nudged, otherwise it is likely to be noisy.)

It is highly likely that the difference in the autoconversion scheme between NSW6 and NDW6 is having a large impact on the response of the LWP to aerosols (which is very different between the simulations), but this is not mentioned. It should be mentioned and discussed in the results and summary.

In summary, before publication I think this paper needs a bit of work to add some more relevant analysis and discussion – in particular a fairer discussion of the performance of the single moment scheme vs double moment for aerosols and a more thorough analysis of the differences in aerosol- cloud interaction forcing (ERFaci).

[A-1] We appreciate your great contributions to improve our manuscript. Your comments and suggestions are very helpful and motivate us to investigate the results more deeply. We think these points are included in the specific points, so we would like to answer each point below.

Through the revision, we modified figures and tables as follows:

- Figures 1, 2, 5, E1 and E2: We changed the color of the zonal averages (NDW6 blue, NSW6 orange, as in the other figures).
- Figures 2 and 5: We replotted the model results in white for grids with missing satellite data. We replotted the zonal averages of the model results by eliminating the grids

with missing satellite data.

- Figure 4: We changed the caption named “references” to “AeroCom”.
- Figure 7: We changed the subtitle named “AOD” to “AOT”.
- Figure 9: We provided IRFari of all and each aerosol for shortwave & longwave at the TOA & the surface under all & clear sky conditions. We changed the caption named “references” to “Kinne19” and “Thorsen21”.
- Figure 10: We removed the results of IRFari because they were shown in Figure 9. Instead, we newly added the results of ERFari and sum of ERFari and ERFaci. We also modified the ERFaci for shortwave to net ERFaci (for both shortwave and longwave).
- Figure 11: We modified ∂AOT to ∂CCN. We added new parameters such as ∂CDNC, ∂CDR, ∂CA, ∂CF, and net ERFaci to further explore ACI.
- Figure 12: We replaced Table 3 in the original manuscript to Figure 12 in the revised manuscript by adding relevant parameters such as ∂CCN, ∂CDNC, ∂CDR, ∂CA, ∂CF, and net ERFaci.
- Figure 13: To explain possible overestimations of the NSW6-simulated Twomey effect, we newly plotted global budgets of the annual averages of the NDW6- and NSW6-simulated Qc (mixing ratio of cloud droplets) and CDNC (cloud droplet number concentrations).
- Table 1 in the original manuscript: We removed it and added a paragraph to explain the HRM and LRM as references in section 2.5 in the revised manuscript.
- Table 1 in the revised manuscript: We simply moved Table 2 in the original manuscript to Table 1 in the revised manuscript.
- Table 2 in the revised manuscript: We showed global and annual mean values of ERFari, ERFaci, and the sum of ERFari and ERFaci for shortwave, longwave, and net radiation under both all-sky and clear-sky conditions.
- Table A1: We newly added the statistical metric to compare results in this study with the references by Goto et al. (2020).
- Table A2: We simply moved Table A1 in the original manuscript to Table A2 in the revised manuscript, with two exceptions. One, we changed “References” to “References from model results”. Second, we changed the SO_2 production value from 67.5 to 67.7.

Please note that some English was corrected in the revised manuscript.

Specific comments

[C1-2] L126 – “although rain does not directly change cloud water in this case” – It’s not clear what you mean here or why this part of the sentence is necessary. Are you saying that there is no scavenging/collection of cloud water by rain? Or that it happens through a calculated microphysical process rate rather than some direct assumption? You’ll have to elaborate or decide whether this part of the sentence is needed.

[C1-3] L126 – “Thus, the impact of scavenged aerosols on cloud water is inevitably overestimated in single- moment bulk cloud microphysics schemes” – it’s not clear what you mean here either. I can see that there may be an overestimate of the number of droplets present for a given amount of aerosol since droplet numbers cannot evolve and reduce over time in a cloud (partial evaporation, coalescence, etc.) in single moment schemes. What you are saying about scavenging is less clear - are you saying that the impact of aerosol changes is likely to be larger in the single moment scheme due to this direct link between CCN and droplet number? And that this would include a larger effect due to removal of aerosol by scavenging? This would need rewriting to say that if so.

[A1-2 & A1-3] Thank you for your comment on this expression. Sorry for confusing you. After we deeply considered this point, we decided to remove this sentence from the revised manuscript.

[C1-4] L191 – "sulfate are assumed to have unimodal particle size distributions" – what are the shapes of the size distributions? Lognormal?

[A1-4] Thanks for your comment. To clarify this, we changed the "unimodal particle size distributions" into "lognormal particle size distributions" (Line 181 in the revised manuscript).

[C1-5] L207 onwards – you describe the length of the simulations and some other details here – but this is in the aerosol section, which makes it hard to find. The information on the model run length, spin-up, etc. should be in its own subsection. Also you need some details on whether they were free running or nudged to meteorology?

[A1-5] Thank you for your suggestion. We agree. We newly add section 2.3 "Experimental conditions" to the revised manuscript to explain the experimental design. It includes the information on the model run length (6 years), spin-up (1 month using 1 year test simulation), external datasets (emission, SST, and sea ice) to run the models. We did not nudge the meteorological fields. This is also reflected on section 2.3. Please see section 2.3 in the revised manuscript.

[C1-6] Plus, the details on the satellite data used for model evaluation should be in its own subsection too and not in the aerosol section.

[A1-6] Thank you for your suggestion. Yes, we agree it. We move this part to new section 2.4 "Observations" in the revised manuscript.

[C1-7] L209 – "which are obtained from the end of the 1-year aerosol online simulations coupled to NSW6." – you will need to explain what these "online simulations" are? What do you mean by "online"?

[A1-7] Thank for your comment. To escape this unclear expression, we removed the term "online" from the revised manuscript. It just means that the aerosol results are simulated by the NICAM coupled to NSW6.

[C1-8] L261 – "However, the seasonal and horizontal biases of the NSW6-simulated LWP are effectively cancelled, and the global and annual mean values of the NSW6-simulated LWP appear closer to the MAC results." – I can't see much evidence for the NSW6 LWP values having compensating seasonal and horizontal biases...

[A1-8] Thank you for your comment on the difference in the simulated LWP between NDW6 and NSW6. We had to explain this more detail in the original manuscript.
In the tropics where the LWP is larger than the other areas, the NDW6-simulated LWP is lower and not closer to the MAC results than the NSW6-simulated LWP. However, as you mentioned in C1-9 and Elsaesser et al. (2017), it should be noted that the MAC results contain regional biases of up to 25%, especially in the tropics. Even with the largest errors, both NDW6- and NSW6-simulated LWP in the tropics are still underestimated compared to the MAC results.
In the zonal distribution in the lower latitudes (30S-0), the NDW6-simulated LWP is lower than the NSW6 results but comparable to the MAC results. In the latitudes over the western Pacific Ocean and Indian Ocean, both NDW6- and NSW6-simulated LWP are lower than the MAC results. However, in the eastern Pacific Ocean and Southern Atlantic Ocean, NDW6-simulated LWP are lower than the NSW6 results but comparable to the MAC results. Therefore, the overestimation of the NSW6-simulated LWP in the eastern Pacific Ocean and Southern Atlantic Ocean effectively cancel the underestimation of the zonal averages in the simulated LWP and unexpectedly provides closer LWP values to the MAC results in terms of the zonal averages. This situation also occurs in the northern hemisphere in the lower latitudes (30N-0). Therefore, in the lower latitudes (30S-30N), the zonal averages of the NSW6-simulated LWP looks closer to the MAC results, but this is attributed from the compensation errors in the regional distribution (we removed the compensation errors in the seasonality from the original manuscript).
In summary, we added these explanations to the revised manuscript to support our statement of the compensation errors in the regional distribution in the NSW6-simulated LWP as follows (Lines 318-329 in the revised manuscript):
"In the tropics where the LWP is larger than the other areas, the NDW6-simulated LWP is lower and not closer to the MAC results than the NSW6-simulated LWP. Notably, the MAC results contain regional biases of up to 25%, especially in the tropics (Elsaesser et al., 2017), but even with the largest errors, the NDW6- and NSW6-simulated LWPs in the tropics are still underestimated compared to the MAC results. In the horizontal distribution over the eastern Pacific Ocean and Southern Atlantic Ocean at lower latitudes (30°S-0), the NDW6-simulated LWP is lower than the NSW6 results but comparable to the MAC results. However, over the western Pacific Ocean and Indian Ocean at the lower latitudes, both NDW6- and NSW6-simulated LWPs are lower than the MAC results. Therefore, the overestimation of the NSW6-simulated LWP in the eastern Pacific Ocean and Southern Atlantic Ocean effectively balanced the underestimation of the zonal averages of the simulated LWP and unexpectedly led to zonal LWP values closer to the MAC results. This situation also occurs in the northern hemisphere at lower latitudes (30°N-0). Therefore, in the lower latitudes (30°S-30°N), the zonal averages of the NSW6-simulated LWP look closer to the MAC results, but this is attributed to the compensation errors in the regional distribution. As a result, the global and annual mean values of the NSW6-simulated LWP appear closer to the MAC results."

[C1-9] L262 – "In terms of the distribution pattern and seasonal cycle, the NDW6-simulated LWP is closer to the MAC result compared to the NSW6 result." – I don't see any evidence for this. Although I would be wary of trusting the MAC LWP values too much in precipitating regions like the tropics since precipitation can cause biases.

[A1-9] Thank you for the useful information about possible bias in the MAC LWP. We checked the reference of Elsaesser et al. (2017) again and found that the statistical error of LWP is large in the tropics by at most 25% regionally. But even when we consider this error, the NICAM-simulated LWP over the tropic is underestimated compared to the MAC result. This point is added to the revised version, as we mentioned in A1-8.

[C1-10] L275 – “the results in NDW6 are generally better and closer to those of the real atmosphere.” – I’m afraid that I don’t think that you can make this statement with the evidence presented. The differences in precipitation (Fig. 1) between NSW6 and NDW6 are very small and it’s hard to say which is best when looking at the spatial maps for January and July. Similarly for LWP. Perhaps NDW6 is looking a bit better for July for the higher latitudes, but it’s quite hard to tell with the colour range chosen – maybe you need to narrow the colour range on the maps to make the differences clearer? But I think it’s clear from the maps and the zonal means that NSW6 does better in the tropics and NDW6 better at higher latitudes even when taking into account the spatial variability and individual months. This would also need addressing at L517.

[A1-10] Thank you for your comment. As you mentioned, the difference in the simulated precipitation between NDW6 and NSW6 is very small. We think the NDW6-simulated LWP is much better than the NSW6-simulated results, but this may be not so clear for readers. Especially, the term “real atmosphere” was removed from our conclusion in the original manuscript because the observation also includes some uncertainty. Therefore, we changed this expression (Lines 340-342 in the revised manuscript) as follows:
“Therefore, the results of precipitation in both NDW6 and NSW6 are comparable to the observations, but those of LWP in NDW6 are different from those in NSW6. The NDW6-simulated LWPs are generally closer to the observations, except for the tropics.”
Thank you for your suggestions to change the color range in Figure 2. As you suggested, we changed them in the revised manuscript.

[C1-11] L288 – “the recent models participating in the AeroCom Phase-III project” – are these what you term “references” in Figure 4? This should be explained in the caption. It would also be better to label this as “Aerocom models” or similar in the figure(s). Similarly, for Fig. 9 – you need separate markers and labels for Kinne (2019) and Thorsen (2021) and you need to explain why there are several crosses for all-sky and clear-sky for “All” and “Anthropogenic”, but not the other plots in the caption – e.g., are they for different AeroCom model members? Are there also several crosses for Thorsen and what are they if so?

[A1-11] Thank you for your suggestions to clarify the references. We modified these labels and captions. In addition, we newly explained these references in section 2.5 named “reference datasets” of the revised manuscript (Lines 296-302) as follows:
“In addition to the results in Goto et al. (2020) as references for a comparison of global aerosol budgets and aerosol optical properties, results obtained from the AeroCom Phase-III project (Gliß et al., 2021) are used in this study. AeroCom Phase-III includes 14 global models and can be the best reference to evaluate global aerosol simulations. For references of the IRFari, the Max Planck Aerosol Climatology version 2 (MACv2 by Kinne, 2019) provides global maps for aerosol optical and radiative properties by

calculating an offline radiative transfer model with the ensemble mean among the AeroCom global models and the in-situ measurements of AERONET. Another reference for IRFari is the mean value from more than 10 studies based on the observations in Thorsen et al. (2021). The IRFari in Thorsen et al. (2021) is only estimated in the shortwave at the TOA."

[C1-12] L310 – "the NDW6-simulated RPCW is much closer to the CloudSat-retrieved RPCW" – again, I have to disagree here. It is true for the subtropical regions and higher latitudes, but not for the tropics.

[A1-12] Thanks for your comment. We modified this in the revised manuscript (Lines 366-369) as follow:
"Because the NSW6-simulated clouds are larger in most regions except for in the tropics, the NDW6-simulated RPCW is much closer to the CloudSat-retrieved RPCW. In the western Pacific Ocean over the tropics where the simulated aerosols are low, the NSW6 results are closer to the CloudSat results.

[C1-13] L320 – "The difference in the column burden of SO2 between NDW6 and NSW6 is caused by the chemical loss in the aqueous phase (0.7 TgS yr-1 or +2%) and gas phase (1.1 TgS yr-1 or -7%) and wet deposition (0.4 TgS yr-1 or +24%), as shown in Table A1." – these values seem different to those quoted in Table A1?

[A1-13] Thank you for your suggestion. We had some mistakes and corrected them in the revised manuscript (Lines 387-389) as follows:
"The difference in the column burden of SO_2 between NDW6 and NSW6 is caused by the chemical loss in the aqueous phase (<u>0.5</u> TgS yr^{-1} or <u>+1</u>%) and gas phase (<u>-1.3</u> TgS yr^{-1} or <u>-10</u>%) and wet deposition (<u>0.5</u> TgS yr^{-1} or <u>+23</u>%), as shown in Table <u>A2</u>."
Because we added a new table (Table A1 in the revised manuscript), Table A1 in the original manuscript was changed to Table A2 in the revised manuscript.

[C1-14] L323 – Need to describe HRM and LRM somewhere in the main paper.

[A1-14] Thank you for your comment on the explanation of HRM and LRM. Yes, we agree. So, we newly added section 2.5 named "reference datasets" to the revised manuscript as follows:
"Our previous model results provided in Goto et al. (2020) using NICAM.16 at a global 14-km high resolution (hereafter referred to as the HRM) and a global 56-km low resolution (hereafter referred to as the LRM) are used as references to compare the NICAM results. As mentioned in section 2.1, the number of vertical layers is set at 38, and the timestep is 1 minute in both the HRM and LRM. The integration periods in both the HRM and LRM are 3 years as climatological runs. The emission inventories, i.e., 2010 for anthropogenic sources, climatological average in 2005-2014 for biomass burning, and natural sources in the present era, and the nudged SST and sea ice in this study are identical to those in both the HRM and LRM, but the initial conditions in this study are different from those in both the HRM and LRM, which use the model results at the end of December after a 1.5-month spin-up. The initial conditions for the model spin-

up are prepared by the reanalysis datasets of the National Centers for Environmental Prediction (NCEP) Final (FNL) (Kalnay et al., 1996) in November 2011. In the cloud microphysics and autoconversion modules, NDW6 coupled to Seifert and Beheng (2006) and NSW6 coupled to Khairoutdinov and Kogan (2000) are used in this study, whereas NSW6 coupled to Berry (1967) is used in both the HRM and LRM. The improvement in the aerosol module described in section 2.2 is also different from that in the HRM and LRM. The results of the HRM and LRM are useful for evaluating the current model results because the observations are limited in some parameters, such as aerosol global budgets and radiative forcings."
Because we created this section, we removed Table 1 in the original manuscript.

[C1-15] L375 – "example, the IRFari dust values are calculated to be -0.46 Wm-2 (NDW6), -0.57 Wm-2 (NSW6), and -0.24 Wm-2 (Kinne, 2019)." – need to make it clear that these are for all-sky.

[A1-15] Thanks. We added "at the TOA under all-sky conditions" in the revised manuscript.

[C1-16] L375 – it would be good to add a bit of detail about what kind of data Kinne (2019) and Thorsen (2021) represent. I.e., model aerosol reanalysis, satellite observations, etc.

[A1-16] Thank you for your comment on the datasets of Kinne (2019) and Thorsen et al. (2021). According to your suggestion, we add the explanation about these references to section 2.5 named "reference datasets" in the revised manuscript. Kinne (2019) provided global maps for aerosol optical and radiative properties by calculating an offline radiative transfer model with the ensemble mean among the AeroCom global models and the in-situ measurements of AERONET. Thorsen et al. (2021) provided the mean value for the shortwave at the TOA from more than 10 studies based on the observations.

[C1-17] L376 – "This is partly caused by the weaker absorption of AOT in this study compared to the median value of the AeroCom models" – But it also could be due to the higher dust AOT values?

[A1-17] Thank you for your suggestion. Yes, we agree. We added the phrase "and the higher dust AOT" in the sentence (Line 445) in the revised manuscript.

[C1-18] L379 – "This is inconsistent with the results of the larger column burden and AOT of dust in this study compared to those of the AeroCom models" – you should point out that it is consistent with too little SW absorption, though.

[A1-18] Thanks. Yes, we agree. We added the sentence "This is consistent with too little shortwave absorption, but" to the revised manuscript (Line 448).

[C1-19] L388 – “For other nonlight-absorbing components, i.e., sea salt and sulfate, the difference in the IRFari values between all-sky and clear sky conditions is very small” – this doesn’t look to be the case? The differences between all-sky and clear-sky are largest for sulphate and sea-salt?

[A1-19] Thanks for your comment on this. We had a mistake and wanted to mention the difference between TOA and surface is the smallest. In the revised manuscript, we modified this in the revised manuscript (Lines 458-459) as follows:
“For other nonlight-absorbing components, i.e., sea salt and sulfate, the difference in the IRFari values between the TOA and the surface is very small.”

[C1-20] L390 – Make it clear that this is for all-sky. Same for L396.

[A1-20] Thanks. We added the word “under all-sky conditions”.

[C1-21] L451 – “The difference in the ERFaci between NDW6 and NSW6 may be partly explained by a nonlinear relationship of the ERFaci to AOT under the different LWPs, as proposed by Carslaw et al. (2013) who argued that even if the aerosol difference between PI and PD is similar, the value of ERFaci can be larger when the aerosol concentration is lower.”

- Presumably you mean here “for different baseline AOT fields”?
- Also “even if the aerosol difference between PI and PD is similar” would be clearer as “even if the PI to PD aerosol difference for two simulations are similar”.

[A1-21] Yes. We wanted to point out that the different baseline of AOT fields can provide the difference in the ERFaci between two experiments, even if the difference in the AOT between two experiments is small. We think your second comment is a clear statement. However, after we carefully analyzed this, we concluded that the difference in the baseline of aerosols between NDW6 and NSW6 did not cause the difference in the ERFaci between NDW6 and NSW6. Surely, Carslaw et al. (2013) and Wilcox et al. (2015) pointed out this possibility of the differences in the ERFaci among the experiments. In this study, the different baseline of AOT between NDW6 and NSW6 under the present days can be found at most 20%, so we thought this can be a reason of the difference in the ERFaci between NDW6 and NSW6 in the original manuscript. However, when we looked at CCN, which is more sensitive to the ERFaci, the different baseline of CCN at 1-km height between NDW6 and NSW6 at the preindustrial days was smaller even in Europe where the difference in the ERFaci between NDW6 and NSW6 was the largest among the regions. Therefore, we modified this in the revised manuscript (Lines 583-588) as follows:
“Carslaw et al. (2013) and Wilcox et al. (2015) pointed out that the different baselines of aerosol fields can provide small differences in ERFaci between two simulations. As mentioned in the previous sections for aerosols, the NDW6-simulated aerosols are generally lower than the NSW6 results, for example IRFari is approximately 15% lower. However, the baseline of CCN at 1-km height between NDW6 and NSW6 under the PI conditions is not very different, so the difference in the baseline of aerosols between NDW6 and NSW6 does not cause the difference in ERFaci between the two simulations.”
We also added the following sentence to the summary (Lines 645-647) in the revised

manuscript:
"Other possible reason for the differences in the ERFaci between NDW6 and NSW6 is the different baselines of aerosol fields, as suggested by Carslaw et al. (2013) and Wilcox et al. (2015), but this is minor because the baseline of CCN at 1-km height between NDW6 and NSW6 under the PI conditions is not very different."
Based on the above discussion, we removed the related statements from the abstract from the original manuscript.

Reference:
Wilcox, L. J., Highwood, E. J., Booth, B. B. B., and Carslaw, K. S.: Quantifying sources of inter-model diversity in the cloud albedo effect, Geophys. Res. Lett., 42, doi:10.1002/2015GL063301, 2015.

[C1-22] L455 – "Figure 11 shows that the horizontal distribution of changes in the simulated LWP" – it would be good to introduce the LWP adjustment as another potential factor in causing the difference in the ERFaci between NDW6 and NSW6 (since you have previously suggested that the baseline AOT is a potential cause).
But what about potential differences in the Twomey effect and cloud lifetime adjustments? How big might these be in NDW6 vs NSW6? Can you compare the PI to PD CDNC differences for NDW6 to the CCN differences in NSW6? Can you compare changes in the cloud fraction (particularly low altitude clouds) – this should be possible if the simulations were nudged. Otherwise it is likely to be noisy.

[A1-22] Thank you very much for your comments on the difference in the ERFaci between NDW6 and NSW6. According to many suggestions of you and other reviewers, we checked other parameters such as CCN, CDNC, CDR (effective radius of clouds), CF (cloud fraction), CA (cloud albedo), and net ERFaci and largely modified our analysis in the revised manuscript (Lines 526-582) as follows:
"Given the verification of the NICAM-simulated CRF above, the simulated ACI due to anthropogenic aerosols is discussed by comparing the results between NDW6 and NSW6 for simulations with aerosol and precursor gas emissions for the preindustrial (PI), mentioned in section 2.3, and the present day (PD). Figure 11 shows the global maps of changes in the simulated CCN at 1-km heights, cloud droplet number concentrations (CDNC) at 1-km heights only for NDW6, cloud droplet effective radius (CDR) at 1-km heights, LWP, cloud albedo (CA), cloud fraction (CF) at 1-km height and net ERFaci between PD and PI. Figure 12 also shows the average values of the selected regions. These figures show that the global average of the NDW6-calculated ∂CCN at a 1-km height is estimated to be 16.70 cm^{-3} (∂CCN), whereas that in NSW6 is estimated to be 19.59 cm^{-3} (∂CCN). The NDW6-calculated ∂CCN values are lower than the NDW6 results. In ∂CDNC, the NDW6-estimated values are +0.70 cm^{-3} (global), +4.22 cm^{-3} (the United States), +4.58 cm^{-3} (Europe), +3.57 cm^{-3} (East Asia), and +0.34 cm^{-3} (India). However, the CDNC used in NSW6 is equal to the CCN concentrations due to the ignorance of sink process in the CDNC in NSW6, as mentioned in section 2.1, so the difference in ∂CDNC between NDW6 and NSW6 is very large. As a result, the NSW6-simulated ∂CDR values at the 1-km height are much larger than the NDW6 results. The NDW6-estimated ∂CDR is -0.17 μm (global), -0.64 μm (the United States), -0.55 μm (Europe), -0.91 μm (East Asia), and -0.33 μm (India), whereas the NSW6-estimated ∂CDR is -0.34 μm (global), -0.93 μm (the United States), -0.91 μm (Europe), -1.20 μm (East Asia), and -

0.81 μm (India). As shown in Figure 11, the NDW6- and NSW6-estimated ∂CDR values are negative near the industrial regions where the ∂CCN is large. Therefore, the approximately 15% difference in ∂CCN between NDW6 and NSW6 causes the approximately 50% difference in ∂CDR. This indicates that the Twomey effect, i.e., the response of ∂CDR to ∂CCN, in NSW6 is larger than that in NDW6.

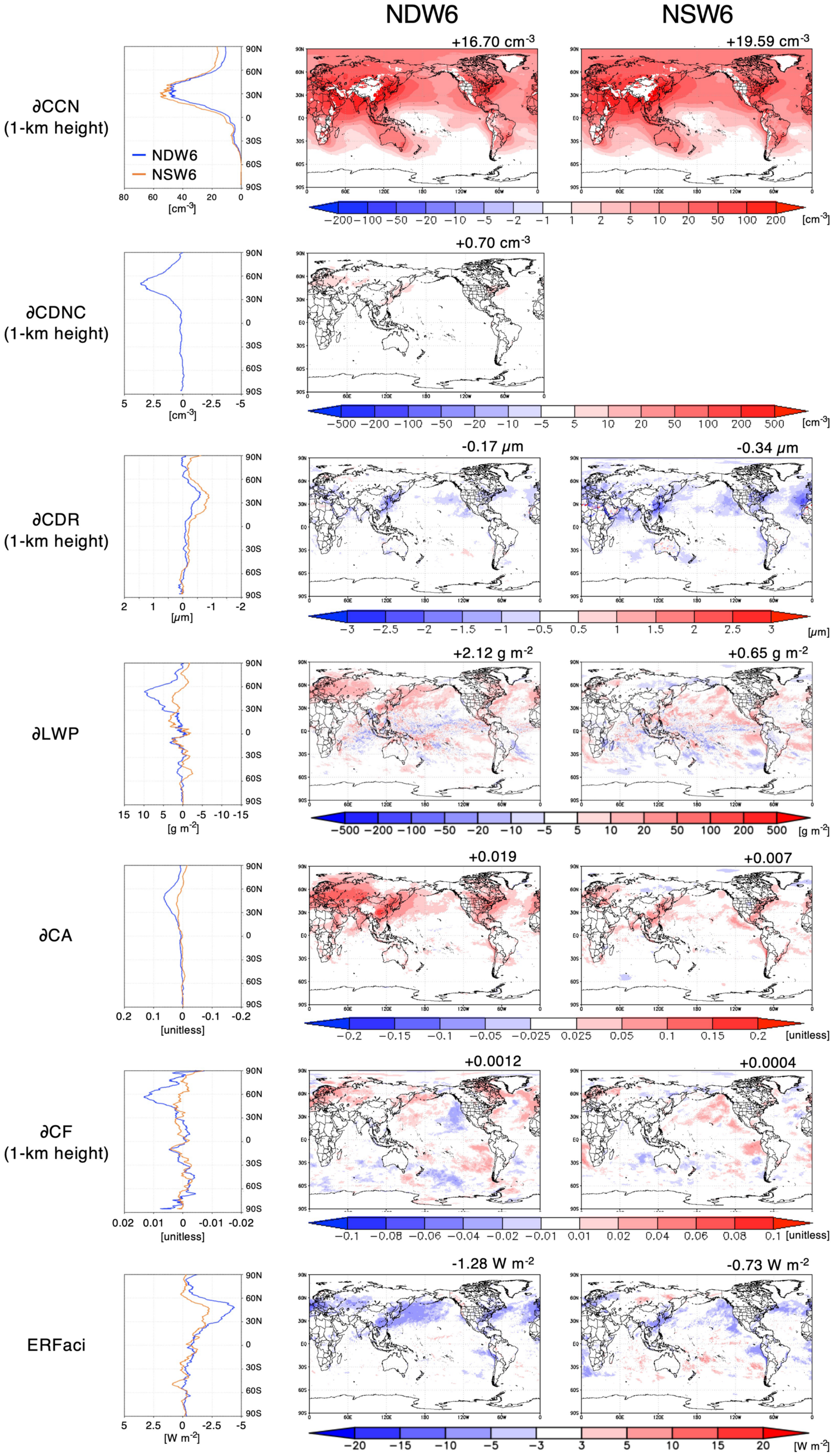
To evaluate the Twomey effect in NDW6 and NSW6, the global averages of differences in the mixing ratios and number concentrations for clouds between the PD and PI aerosol conditions are plotted in Figure 13. The changes in ∂Qc in both NDW6 and NSW6 are positive at most heights, so Qc increases as aerosols increase. This is consistent with the results of ∂LWP shown in Figures 11 and 12(e). The largest value of ∂Qc in both NDW6 and NSW6 occurs at a height of approximately 1.5 km, but the largest values in NDW6 are distributed up to a height of 2 km. Above a height of 3 km, ∂Qc in NDW6 is positive, whereas ∂Qc in NSQ6 is close to zero or negative. This difference in ∂Qc between NDW6 and NSW6 is possibly caused by the differences in the simulated supercooled liquid water in mixed-phase clouds, as mentioned in section 3.1. For ∂CDNC, the largest values in NDW6 occur at a height of 1.2 km, which is slightly lower than the height where the largest value of ∂Qc occurs. This reflects the vertical structure of typical clouds in NDW6. In contrast, the vertical profile of ∂CDNC in NSW6 is different from that of ∂Qc because NSW6 cannot predict CDNC and adopts ∂CCN. This implies that ∂CDR is anti-proportional to ∂Qc from the surface to the 4-km height and has a low value below the 1.5-km height and the largest value at a height of approximately 1.5 km. Specifically, above a height of 3 km, where ∂Qc is close to zero and ∂CDNC has a positive value, ∂CDR should be small. The possible overestimation of ∂CDR in NSW6 represents possible overestimation of the Twomey effect in NSW6.

+4.96 g m^{-2} (the United States), +2.52 g m^{-2} (Europe), +2.62 g m^{-2} (East Asia), and -0.44

As mentioned above, the NDW6-calculated ∂LWP values are higher than the NSW6 results by three times in global averages. The NDW6-estimated values are +2.12 g m^{-2} (global), +7.52 g m^{-2} (the United States), +15.45 g m^{-2} (Europe), +8.77 g m^{-2} (East Asia), and +3.36 g m^{-2} (India), whereas the NSW6-estimated values are +0.65 g m^{-2} (global), +4.96 g m^{-2} (the United States), +2.52 g m^{-2} (Europe), +2.62 g m^{-2} (East Asia), and -0.44 g m^{-2} (India). The positive values in ∂LWP in both NDW6 and NSW6 could be caused by a decrease in auto-conversion due to the increase in CDNC. However, magnitudes of ∂LWP differ between NDW6 and NSW6, which is the largest in Europe among others, whereas the NDW6- and NSW6-simulated ∂CCN are close to each other in most regions. This appears to indicate that the cloud water susceptibility, defined as the difference in ∂LWP against ∂CCN from PD to PI conditions, is larger in NDW6 than in NSW6. Such a different susceptibility could be interpreted in terms of different complexities of hydrometeors interactions between NSW6 and NDW6, particularly whether or not the CDNC and rain drop number concentration (RDNC) are predicted. This generates different variabilities of CDNC and RDNC between the two schemes, possibly leading to the different susceptibilities. Nevertheless, more detailed analysis will be required in future studies to explore microphysical processes responsible for these different behaviors between the two schemes.

The horizontal distribution of changes in the simulated ERFaci is generally consistent with changes in the simulated ∂LWP (Figure 11). By decreasing the simulated ∂CDR, increasing the simulated ∂LWP from PI to PD, and increasing the simulated ∂CA and ∂CF at 1-km height, the negative values of the simulated ERFaci in industrial regions, such as the United States, Europe, and East Asia, increase in magnitude. The global annual averages of the net ERFaci value are estimated to be -1.28 Wm^{-2} (NDW6) and -

0.73 Wm^{-2} (NSW6). Both NDW6- and NSW6-estimated ERFaci values range within the results in IPCC-AR6 (Forster et al., 2021), i.e., -0.84 Wm^{-2} (-1.45 Wm^{-2} to -0.25 Wm^{-2}), and the Radiative Forcing Model Intercomparison Project (RFMIP) (Smith et al., 2020), i.e., -0.81±0.30 Wm^{-2}, by considering the uncertainty caused by the assumption in the PI conditions. The magnitude of the ERFaci value in NDW6 is larger than that in NSW6 by 0.55 Wm^{-2} (approximately 43% of the ERFaci value in NDW6), whereas the NDW6-simulated aerosol loadings are smaller than the NSW6 results, as shown in the previous sections. Figure 12 shows that the negative NDW6-estimated ERFaci values are larger than the NSW6-estimated ERFaci values by 2.33 Wm^{-2} (US), 3.22 Wm^{-2} (Europe), 1.10 Wm^{-2} (East Asia), and 0.89 Wm^{-2} (India). Therefore, it was suggested that the ERFaci due to the cloud lifetime effect in NDW6 was larger than that in NSW6 due to the Twomey effect, although the NSW6-simulated ERFaci certainly includes some bias due to the overestimation of the Twomey effect."



NDW6
NSW6
∂CCN (1-km height)
NDW6
NSW6
[cm-3]
+16.70 cm-3
+19.59 cm-3
−200 −100 −50 −20 −10 −5 −2 −1 1 2 5 10 20 50 100 200 [cm-3]
∂CDNC (1-km height)
[cm-3]
+0.70 cm-3
−500 −200 −100 −50 −20 −10 −5 5 10 20 50 100 200 500 [cm-3]
∂CDR (1-km height)
[μm]
-0.17 μm
-0.34 μm
−3 −2.5 −2 −1.5 −1 −0.5 0.5 1 1.5 2 2.5 3 [μm]
∂LWP
[g m-2]
+2.12 g m-2
+0.65 g m-2
−500 −200 −100 −50 −20 −10 −5 5 10 20 50 100 200 500 [g m-2]
∂CA
[unitless]
+0.019
+0.007
−0.2 −0.15 −0.1 −0.05 −0.025 0.025 0.05 0.1 0.15 0.2 [unitless]
∂CF (1-km height)
[unitless]
+0.0012
+0.0004
−0.1 −0.08 −0.06 −0.04 −0.02 −0.01 0.01 0.02 0.04 0.06 0.08 0.1 [unitless]
ERFaci
[W m-2]
-1.28 W m-2
-0.73 W m-2
−20 −15 −10 −5 −3 3 5 10 15 20 [W m-2]

Figure 11: Global distributions of the annual averages of the NDW6- and NSW6-simulated CCN change at 1-km height (∂CCN), Nc (cloud droplet number concentrations at 1-km height) change (∂CDNC), CDR (cloud droplet effective radius for warm clouds at 1-km height) change (∂CDR), LWP change (∂LWP), CA (cloud albedo) change (∂CA), CF (cloud fraction at 1-km height) change (∂CF), and net ERFaci by comparing the results between NDW6 and NSW6 for simulations with aerosol and precursor gas emissions for the present and the preindustrial era. The number located in the upper right in each panel represents the global and annual mean value. The results at 1-km height also include areas with elevations higher than 1-km height in white.

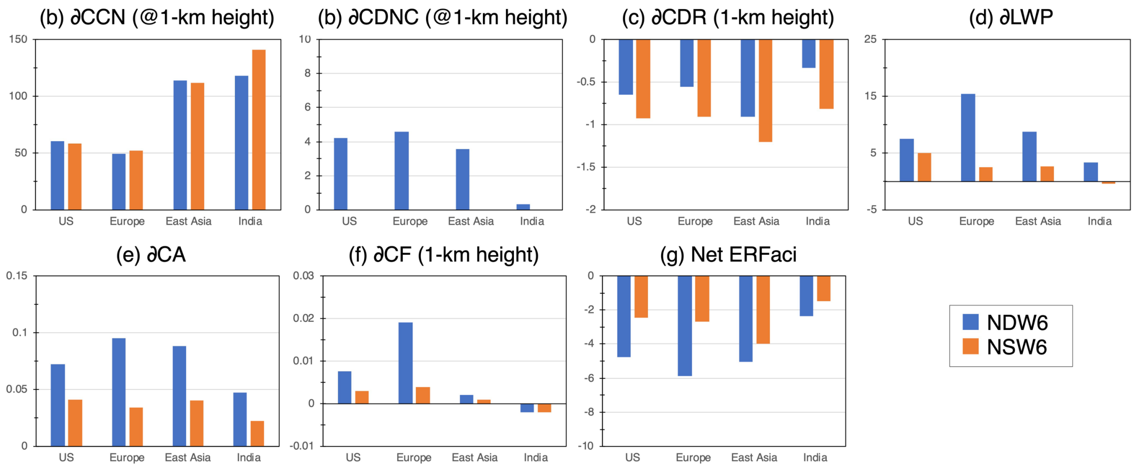


Figure 12: Regional averages of the differences in CCN at 1-km height, CDNC (cloud droplet number concentration only in NDW6), CDR (cloud droplet effective radius at 1-km height), LWP, CA (cloud albedo), CF (cloud fraction at a 1-km height), and net ERFaci between the preindustrial and the present days. The regions are defined as US (90°W-60°W, 30°N-50°N), Europe (0°E-30°E, 40°N-60°N), East Asia (110°E-140°E, 20°N-50°N), and India (70°E-90°E, 10°N-35°N).

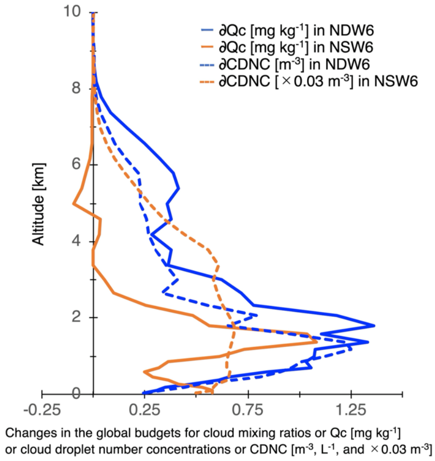


Figure 13: Global budgets of the annual averages of the NDW6- and NSW6-simulated Qc (mixing ratio of cloud droplets), the NDW6-simulated CDNC (cloud droplet number concentration), and the NSW6-simulated CDNC (cloud droplet number concentration, which is equal to CCN number concentrations)

[C1-23] L466 – "Notably, the reason for the differences in the susceptibility between NDW6 and NSW6 should be addressed in future studies."

- Although it is clear that the difference in the autoconversion scheme between NSW6 and NDW6 is likely to have a large impact here and so this should be mentioned and discussed here and in the summary too.

[A1-23] Thank you for your comment on the autoconversion scheme. Theoretically, the dependence of the autoconversion on CDNC in Seifert and Beheng (2006) (hereafter referred to as SB06) is larger than in Khairoutdinov and Kogan (2000) (hereafter referred to as KK00), so for the same ∂CDNC, SB06 has a lower cloud-to-rain conversion efficiency (clouds in SB06 tend to remain). Because NDW6 uses SB06 and NSW6 uses KK00, if we look only at the autoconversion, NDW6 has a larger ∂LWP than NSW6. This tendency is consistent with the results of the difference in ∂LWP between NDW6 and NSW6, as shown in Figure 11 in this study, but the magnitude of the dependence on CDNC is not very large between SB06 and KK00. Therefore, we think that the difference in the autoconversion between SB06 and KK00 alone cannot explain the difference in the ∂LWP between NDW6 and NSW6 in this study.
Ideally, we should do sensitivity experiments for different autoconversion schemes, but we cannot do them due to the limitation of available computer resources in our environment. So, we would like to discuss possible differences in the ERFaci among different autoconversion schemes by using a reference. Michibata and Suzuki (2020) shows the difference in the ERFaci between various autoconversion schemes on the MIROC climate model. This version of MIROC adopts a double-moment cloud microphysics module including prognostic precipitation. The sensitivity experiments in the MIROC used various autoconversion including KK00 and SB06. Therefore, Michibata and Suzuki (2020) can be a suitable reference here. Even though Michibata and Suzuki (2020) did not deeply discuss the difference in the ERFaci among the different autoconversion schemes, the difference in the global and annual mean values of ERFaci between KK00 and SB06 can be estimated to be 0.15 Wm^{-2}. The magnitude of the ERFaci with SB06 is smaller than that with KK00, although the dependence of the autoconversion on CDNC in SB06 is larger than in KK00. This suggests that the difference in the LWP response to aerosol between SB06 and KK00 is not large and that the difference in the ERFaci may have opposite signs through processes other than autoconversion. Therefore, we added the following sentences to the revised manuscript (Lines 589-594):
"The difference in the autoconversion from clouds to precipitation between NDW6 and NSW6 can be a reason for the difference in ERFaci between NDW6 and NSW6. Using a global aerosol model, MIROC, coupled to a double-moment bulk cloud microphysics scheme with coarse resolution of 1.4° × 1.4°, the difference in ERFaci between Khairoutdinov and Kogan (2000) and Seifert and Beheng (2006) is estimated to be 0.15 Wm^{-2} (Michibata and Suzuki, 2020). This magnitude of ERFaci difference potentially caused by the two different autoconversion schemes cannot explain the difference in ERFaci between NDW6 and NSW6 of this study."
We also added the following comments to the summary (Lines 647-652) in the revised manuscript:
"Another possible reason is the difference in the autoconversion between NDW6 and NSW6, and the difference in ERFaci between Khairoutdinov and Kogan (2000) and Seifert and Beheng (2006) is estimated to be 0.15 Wm^{-2} by using a general circulation

model MIROC (Michibata and Suzuki, 2020). However, this magnitude of ERFaci difference potentially caused by the two different autoconversion schemes cannot explain the difference in ERFaci between NDW6 and NSW6 of this study."

Reference:

Michibata, T., and Suzuki, K.: Reconciling compensating errors between precipitation constraints and the energy budget in a climate model, Geophys Res. Lett., 47, e2020GL088340, doi:10.1029/2020GL088340, 2020.

[C1-24] L496 – "The differences in the dust emissions, dust column burden and SO2, AOT, and IRFari values for total aerosols between NDW6 and NSW6 are larger." – at L480 you said that the AOT differences were small, so this is a bit of a contradiction.

[A1-24] Thanks for your comment on this remark. We added the term "than those in the other aerosol budgets and components" in the revised manuscript (Line 623).

[C1-25] L508 – "These differences are mainly caused by the difference in the susceptibility of the LWP to AOT" – in the results section this wasn't stated so clearly. It would be good to mention that in the results too – although do you have evidence that this is the case?

[A1-25] Thank you for your comment. In A1-22, we drastically modified the explanation about the ERFaci by showing the relevant parameters, such as CCN, CDNC, CDR and LWP. One of the important conclusions is that the ERFaci due to the cloud lifetime effect in NDW6 was larger than that in NSW6 due to the Twomey effect. The difference in the cloud lifetime effect between NDW6 and NSW6 is possibly caused by the difference in the treatment of cloud droplet and raindrop number concentrations. Unlike NSW6, NDW6 predicts both CDNC and raindrop number concentration (RDNC). This generates different variabilities of CDNC and RDNC between the two schemes, possibly leading to the different susceptibilities. Nevertheless, more detailed analysis will be required in future studies to explore microphysical processes responsible for these different behaviors between the two schemes.

Another important conclusion is that the NSW6-simulated ERFaci certainly includes some bias due to the overestimation of the Twomey effect. This is clearly shown in Figure 13, which shows that the vertical profile of ∂CDNC in NSW6 is different from that of ∂Qc because NSW6 cannot predict CDNC and adopts ∂CCN. Therefore, we modified this sentence in the summary (Lines 637-645) in the revised manuscript as follows:

"These differences are mainly caused by the difference in the susceptibility of the ∂LWP to ∂CCN. <u>As discussed in section 4.2, it was suggested the increase in changes in ERFaci due to the cloud lifetime effect in NDW6 is larger than that in NSW6 due to the Twomey effect, although the NSW6-simulated ERFaci certainly includes some bias due to the overestimation of the Twomey effect. The different susceptibility between NDW6 and NSW6 could be interpreted in terms of different complexities of hydrometeors interactions between NSW6 and NDW6, particularly whether the CDNC and RDNC are predicted. This generates different variabilities of CDNC and RDNC between the two schemes, possibly leading to the different susceptibilities. Nevertheless, more detailed analysis will be required in future studies to explore microphysical processes responsible for these different behaviors between the two schemes.</u>

[C1-26] L518 – "and thus, the use of NDW6 is recommended in environmental and climate simulations." - The justification regarding the LWP and precipitation performance here is not very strong (see above regarding the NSW6 vs NDW6 performance). Perhaps this would be better justified by the likely better representation of cloud microphysical processes in NDW6 (e.g, a better representation of fall speeds, hydrometeor separation, droplet evaporation, etc.?). And the large difference in ERFaci are also a good reason to consider NDW6 - although which one is more accurate is unknown... It might be worth mentioning and considering that the LWP response to aerosol could be constrained/evaluated to some degree with a short simulation of a well-observed volcanic eruption, e.g., like the Holuhraun eruption (Malavelle, Nature, 2017; doi:10.1038/nature22974)

[A1-26] Thank you for your comment on the recommendation to use the NDW6 scheme. As you said that "this would be better justified by the likely better representation of cloud microphysical processes in NDW6", we think this should be reflected in the revised manuscript. Our answer using Figures 11,12, and 13 in the revised manuscript showed that the differences in the ∂CCN and ∂CDNC between NDW6 and NSW6 were large, and thus the differences in the ∂CDR between NDW6 and NSW6 were large. Therefore, from this point as well, it is considered that the reliability of the calculation results using NDW6 is high, and in the revised manuscript (Lines 667-669), we have corrected this as follows:
"The cloud microphysics representation of NDW6 is more elaborate than that of NSW6, and it found that the NSW6-simulated CDR is overestimated due to the inability to predict CDNC in NSW6. Therefore, the use of NDW6 is recommended in environmental and climate simulations".
And because we found a clear bias of the results in NSW6, we also removed the following part from the end of the paragraph in the revised manuscript: "However, because simulations using NDW6 require 1.5 times more calculation resources, the use of NSW6 is still useful for long-period climate simulations at high resolutions."
Thank you for your suggestion to check the literature on numerical experiments for the Holuhraun eruption. It can be a reference to justify the response of aerosols to clouds. According to Malavelle et al. (2017), the CDR decreases but the LWP doesn't clearly increase or decrease when the CCNs increases. In the revised Figure 11, it clearly shows that the NDW6- and NSW6-simulated CDRs decrease and the NDW6- and NSW6-simulated CDRs increase when the CCN increase. The CDR response to aerosol seems to be consistent to the results in Malavelle et al. (2017), but the LWP response to aerosol seems to be different from the results in Malavelle et al. (2017). Even if we perform the volcanic eruption experiment, we can only evaluate a specific case. In addition, this area (60-70N) and this period (September-October) may be affected not only by interactions between aerosols and clouds, but also by interactions between cloud water and cloud ice. As you mentioned "to some degree", we agree this point and think that it is still difficult to evaluate the LWP response to aerosols in this study. Therefore, although we did not perform this extra experiment in this study, we would like to do so in future studies to investigate the LWP response to aerosol in one case. Therefore, we added this following sentence to the end of conclusion in the revised manuscript (Lines 669-671):
"At the same time, because the ERFaci in NDW6 needs validation, in the future it will be necessary to perform additional experiments targeting specific cases of volcano in

Iceland shown in Malavelle et al. (2017) to deeply evaluate the model results in NDW6."

[C1-27] L529-531 – I can't see the PCC, etc. values from Goto (2020) listed anywhere – these should be included or quoted in this paper somewhere – ideally in a table so that the reader can compare the new simulations to the old ones.

[A1-27] Thank you for your comment on these values and suggestion. To clarify them, we added a new table (Table A1 in the revised manuscript) to summarize the statistical metrics. At least, these values are our mistake, so they are modified in the revised manuscript (Line 678) as follows:
"(e.g., PCC of 0.819, RMSE of 5.03, and NMB of -54.8 μg m^{-3} from Goto et al., 2020)".

Table A1 Statistical metrics of PCC, RMSE, and NMB for the annual averages of surface aerosol mass concentrations (OM, BC, and sulfate) between in situ measurements and the NICAM simulations (NDW6 and NSW6 in this study shown in the panels of Figure 3 and HRM and LRM in Figure 8 in Goto et al., 2020).

	NDW6	NSW6	HRM	LRM
		OM		
PCC	0.847	0.846	0.819	0.794
RMSE [μg m^{-3}]	3.40	3.34	5.03	5.21
NMB [%]	-30.4	-25.8	-54.8	-56.1
		BC		
PCC	0.904	0.904	0.890	0.869
RMSE [μg m^{-3}]	1.05	1.03	1.16	1.28
NMB [%]	-53.4	-51.3	-46.4	-52.3
		Sulfate		
PCC	0.807	0.853	0.815	0.768
RMSE [μg m^{-3}]	3.97	3.67	3.94	4.34
NMB [%]	-10.4	-3.7	-14.6	-23.7

[C1-28] Table A1 – it would be useful to say what kind of observations the difference references use (satellite, model reanalysis, models, etc.).

[A1-28] Thanks for your comment. They are all model results. We modified "References from model results" in Table A2 of the revised manuscript.

[C1-29] L554 – "whereas the difference in the cloud microphysics module does not affect the chemical budget of SO2 oxidation" – presumably you are talking about the cloud microphysical changes within the NSW6 module here rather than the difference between double and single moment (since there are large differences for SO2 between NSW6 and NDW6)? You should make this clearer if so.

[A1-29] Thank you very much for reading the appendix carefully. We had a mistake. This part (comment on the difference among the different cloud microphysics modules) should not be mentioned in this section, because this part focused on NSW6, HRM, and LRM using NSW6. We removed this in the revised manuscript.

Typos / grammar

[C1-30]

L18 – “but some aerosol species, especially dust and sulfate, have larger differences among the experiments with NDW6 and NSW6 compared to those among the experiments with different horizontal resolutions, i.e., 14 km and 56 km grid spacing,...”
- Would be better as “but differences between the NDW6 and NSW6 experiments are larger for some aerosol species, especially dust and sulfate, compared to those between experiments with different horizontal resolutions, i.e., 14 km and 56 km grid spacings,...”

L48 – “and hence, an elaboration of both the cloud module and aerosol physics module is required to improve ACI in climate models”. “Elaboration” is perhaps not the best choice of words. I recommend “evaluation” instead.

L52 – “It is promising that convective cloud systems are better represented with finer model resolution when cumulus parameterizations are avoided” – I would recommend this instead: “These results suggest that convective cloud systems are better represented with a finer model resolution for which cumulus parameterizations are avoided”

L72 – “solved” -> “resolved”

L88 – “calculcated” -> “run”

L96 - “incorporated to” -> “incorporated into”

L97 – “were reflected to the version in NICAM.19” -> “was incorporated into the version in NICAM.19”

L107 – “to aerosol physics module” -> “to the aerosol physics module”

L157 – “concentration higher” -> “concentration is higher”

L167 – “dependence of leaf area index” – should be “dependence on leaf area index”?

L175 – “OM” has not been defined.

L176 – “Secondary organic aerosols (SOAs) are assumed to be particles by multiplying the emission fluxes of isoprene and terpenes provided by” -> “Secondary organic aerosols (SOAs) are assumed to form particles, which are calculated by multiplying the emission fluxes of isoprene and terpenes provided by” – also, what are they multiplied by? A constant factor?

L178 – “Parts of SO2 are emitted from volcanic eruptions (Diehl et al., 2012) and are formed from DMS,” – “SO2 is emitted from volcanic eruptions (Diehl et al., 2012) and is also formed from DMS,”

L217 – “These identical datasets were prepared and used in Goto et al. (2020)” – I assume you mean “The same datasets were prepared and used in Goto et al. (2020)”. Otherwise it makes it sound like all of the observational datasets are identical.

L249 – “by satellite results” -> “using satellite data”.

L268 – “CF at the low level" -> "low-altitude CF”

L306 – need to define what WSBC and WIBC are somewhere.

L394 – “which can be caused by its lower lifetime among the references” -> “which may be due to its short lifetime relative to the values from Kinne (2019) (and Thorsen?)”.

L395 – “sea salt is more scavenged by wet deposition” -> “sea salt is scavenged more by wet deposition”

L418 – “but comparable to” -> “but are comparable to”

L422 – “are smaller than”-> “are smaller in magnitude than”

L427 – “fluxes are compared for model evaluations of radiation budget.” -> “fluxes are compared and evaluated.”

L427 – “the global and January averages of the SWCRF” -> “the global averages of the SWCRF for January”. Similarly, for L428 for July and L436 for global averages.
L442 – “by comparing the results between NDW6 and NSW6 under the preindustrial (PI) and the present day (PD)” – better to be clear that this means PI and PD emissions for aerosols and precursor gases. E.g., “by comparing the results between NDW6 and NSW6 for simulations with preindustrial (PI) and the present day (PD) aerosol and precursor gas emissions”. Similarly for the caption of Fig. 11.
L457 – “such as the United States, Europe, and East Asia, increase” -> “such as the United States, Europe, and East Asia, increase in magnitude”
L457 – “the NDW6-estimated ERFaci value is larger negatively than the NSW6-estimated ERFaci” -> “the negative NDW6-estimated ERFaci values are larger in magnitude than the NSW6-estimated ERFaci values”
L460 – “key to understand the difference” -> “key to understanding the difference”
L502 – “whereas those in the sulfate are mainly caused by the wet deposition of SO2.” -> “whereas those in the sulfate are mainly caused by the differences in the wet deposition of SO2.”
L530 – “or” -> “of”
L538 – “Surely,” – wrong choice of phrase here. Would be better as “In support of this,”
Figure E1 – the caption has become jumbled.

[A1-30] Thank you very much for reading and checking the details in our manuscript and giving your corrections. We reflected them in the revised manuscript.