Response to Anonymous Reviewer-1

I think this manuscript is within the scope of GMD and does constitute a contribution to the field. The authors showcase (briefly) some recent improvements in an atmosphere model. They then proceed to applying some statistical analysis on the improved model.

The manuscript is decently written, but can be challenging to read. Most figures can be improved. The manuscript covers two distinct topics that at times come across as unrelated. Perhaps this could be two separate manuscripts focused on two separate topics? The model evaluation part is brief and incomplete; likewise, the maximum covariance part only touches the surface of the topic it is tackling.

I have written down some reactions, comments, and questions. I hope the authors find them helpful in revising their manuscript (if they choose to do so). I don't see any major issue in this manuscript and so I will be happy if it is published in the near future :)

We thank the reviewer for the encouraging comments and constructive suggestions. Please find below a point by point response to the queries raised.

The decision to consolidate the two topics into a single manuscript was based on the following:

1. The model evaluation section focuses exclusively on cloud properties, although the model improvements targeted both aerosols and clouds so as to reduce the uncertainties in estimating the radiative forcing exerted by aerosols.
2. The maximum covariance analysis technique employed in this study enables an in-depth exploration of the intricate interplay between aerosols and clouds. This analysis is particularly relevant in three distinct regions where low-level clouds and pollution outflows coincide and are prominent, providing valuable insights into their complex interactions.

By consolidating these elements into one manuscript, we aim to demonstrate how the recent improvements enable us to provide a comprehensive perspective on the interrelated nature of aerosols and clouds.

We have made an in-depth evaluation of all important cloud physical properties that we can obtain from satellites in the manuscript. We have further clarified this in Subsection 3 as 'Assessing the cloud properties with the improved ECE3-FORCeS model'. Though the maximum covariance analysis conducted in this study is at its present state self-contained, we agree that there is the possibility to explore more. We also primarily wanted to introduce and demonstrate the efficacy of this method in extracting information when two variables presumably covary.

Lines 63--69: Not sure what this paragraph adds to the conversation here. Consider adding more context why this is needed or simply deleting.

We agree with the reviewer as this information has already been given in the 'Acknowledgements'. This paragraph has been rewritten as follows ' Within the framework of the FORCES project (https://forces-project.eu/), this study aims to identify and implement the crucial aerosol and cloud processes that could potentially reduce uncertainties in aerosol forcing estimates. This is detailed in Section 2.2. These improvements are being incorporated into the three global models that were previously used for CMIP6 (https://pcmdi.llnl.gov/CMIP6/): NorESM2-LM (Seland et al., 2020), MPI-ESM1.2-HAM (Mauritsen et al., 2019; Neubauer et al., 2019) and EC-Earth3-AerChem (van Noije et al., 2021). This paper focuses on the simulations carried out with the EC-Earth3-AerChem model.'

Lines 76--77: This may warrant further explanation. As a somewhat informed reader, this left me wondering: What's the effective resolution of this model in sum? At what resolution are things like macrophysics and microphysics schemes are done? It is fine to refer to citations; segueing into this with a "however" left me wondering about several questions. So, I would either consider expanding slightly here (e.g., why you felt the need to point out this specific piece of information) or simply delete to avoid reading getting stuck like I was.

The chemistry and aerosol module, TM5, is computationally expensive and is run at a much coarser resolution compared to the atmospheric model. The cloud macrophysics and microphysics are part of the atmosphere model and so are run at the same T255L91 resolution. These resolution details are typically provided when describing the model configuration to facilitate tracking of the specific resolutions employed for a particular evaluation or analysis. This has now been clarified in the revised manuscript.

Line 84: Why brief? Isn't this manuscript supposed to detail these updates?

Here, by 'brief', we meant that we will not be detailing the impacts of all individual improvements (Section 2.2.1 - Section 2.2.5) with respect to the CMIP6 model version. This will be published separately. Here, we are focussing mainly on evaluating the resulting improvements in cloud properties after these updates.

Lines 85--94: Since the Patridge et al. manuscript is in preparation, I worry this entire paragraph (section) lacks substantiation. On the one hand, someone like me having read the studies cited a few times in the past can nod in agreement about the assertion that the 2014 scheme is likely better. On the other hand, how can you convince others that your assertions are sound besides taking your words for it? A potential avenue is weakening the assertions and citing counterarguments (e.g., the meta-modeling study in doi:10.1175/JAS-D-15-0223.1 can be interpreted as saying these models perform similarly an older comparison study doi:10.1029/2011MS000074 show differences between schemes).

We agree with the reviewer and this section has been revised as follows:

The formation of cloud droplets from ambient aerosol particles is a crucial process that needs to be accurately represented in global models to predict aerosol-cloud interactions and hence, the indirect effects of aerosols on climate. To achieve this, the modeling community

has developed parameterization schemes that can realistically simulate the cloud droplet number concentrations based on ambient supersaturation, aerosol size and number distribution and updraft velocities. The schemes range from simpler schemes such as Abdul-Razzak and Ghan (2000) and Shipway (2015) to more complex schemes such as Barahona et al., (2010) and Morales-Betancourt and Nenes scheme (2014). All parameterizations show a good overall agreement with the parcel model, but exhibit different sensitivities to aerosol perturbations under different meteorological conditions or pollution regimes (Ghan et al., 2011). For example, the Abdul-Razzak and Ghan (2000) scheme simulates low/high maximum supersaturations (S_{max}) and activation at low/high aerosol loads compared to the Morales-Betancourt and Nenes (2014) scheme making it more sensitive to increases in aerosol loads. The Shipway (2015) scheme is closer to the Morales-Betancourt and Nenes (2014) scheme, but tends to overestimate S_{max} and activation in polluted environments. Also, the simpler schemes show larger biases compared to the complex schemes in clean marine regions, particularly over the Southern Ocean. Hence, the current cloud activation scheme of Abdul-Razzak and Ghan (2000) is replaced by the Morales-Betancourt and Nenes (2014) scheme in the FORCeS version of the EC-Earth3-AerChem model. This scheme makes use of the population splitting concept wherein the growing population of droplets is divided into three enabling a more accurate estimate for the number concentration of droplets formed at the time of S_{max}.

Lines 95--109: This section may benefit from rewriting and clarifying. As an example, the second sentence seems incomplete, a sentence has a misplaced comma interrupting the flow and meaning of it, and finally the citation style is inconsistent (years sometimes are surrounded by parentheses; other times they are not). More substantially, this section left me wondering about dust aerosol. How is it represented and emitted? How do these changes impact it?

EC-Earth3-AerChem calculates dust emission online (van Noije et al., 2021). The emission scheme considers dry paleolakes as preferential dust sources and accounts for seasonal variations in vegetation cover. Saltation of dust particles occurs when the surface wind exceeds a certain threshold velocity, which depends on the soil particle size distribution and soil conditions. Furthermore, snow cover prevents dust emission in the model (Tegen et al., 2002; 2004). Dust is initially emitted into the accumulation and coarse insoluble modes, defined in the M7 aerosol microphysics module (Vignati et al., 2004). Within FORCeS, the dust emission module has been further developed to explicitly incorporate the atmospheric cycle of dust minerals relevant for their climate impacts. These include: (1) iron oxides, which control the absorption of shortwave (SW) radiation by dust aerosols (e.g., Sokolik and Toon 1999; Di Biagio et al., 2019), (2) quartz and feldspars, which constitute efficient ice-nucleators (e.g., Atkinson et al., 2013; Harrison et al., 2019), and (3) calcite, which intervenes in atmospheric chemistry processes and affects the aerosol pH.

To estimate the emission of the different minerals, we derive the size-distributed mass fraction at emission for each of them considering information on the soil mineralogy of the dust sources (Claquin et al., 1999; Nickovic et al., 2012). To account for the differences in the mineral's size distribution reported in the soil and those found in the aerosols, we apply an extension of the Brittle Fragmentation Theory for dust emission proposed by Kok (2011).

This approach has shown a better agreement with observations (e.g., Perlwitz et al., 2015a, b) than relying exclusively on the soil size fractions. These newly incorporated minerals are then coupled to different processes in the model.

The effect of iron oxides in the SW optical properties for dust is accounted for by assuming internal mixtures of iron oxides and other host minerals. A Maxwell Garnett mixing rule is applied to derive online the dust refractive index, considering the volume fraction of hematite and a homogeneous matrix of other minerals. The reference refractive indices are taken from Scanza et al. (2015).

The abundance of quartz and K-feldspars is considered to derive the ice nucleating particles, and then coupled to the formation of mixed-phase clouds in the model. In order to estimate the specific fraction of K-feldspar, we assume it to be a 35% of the total feldspar originally provided in the soil mineralogy datasets (Atkinson et al., 2013; Chatziparaschos et al., 2023). Finally, the calcite in dust is explicitly linked to the thermodynamic equilibrium of inorganic species, solved in the model with the ISORROPIA-Lite (Kakavas et al., 2022) model (see section 2.2.5).

Line 122: Perhaps use a more conventional notation for the "approximately bigger than" unless you mean something different.

The sentence now reads as 'In regions such as the remote Arctic/Antarctica where the primary ice nucleating particles (INP) are sparse (Moore et al., 2024; Wex et al., 2019), a considerable source of ice crystals is via secondary ice production (SIP) processes in supercooled clouds *(temperatures higher than approximately -10ºC)* (Järvinen et al., 2022; Järvinen et al., 2023).

Lines 134--136: Any reasoning for this choice to parameterize this process?

The notion of the Ice Enhancement Factor (IEF) is frequently employed in the literature to indicate the prevalence of secondary ice production (SIP). This is a promising approach for parameterizing the effects of SIP in mixed-phase clouds because it allows direct comparison with the IEF (ICNCs/INPs) derived from in-situ observations or retrieved from remote-sensing data (e.g., Wieder et al., 2022). Unlike conventional SIP parameterizations (e.g., Phillips et al., 2017), the RaFSIP scheme requires only a few input variables, offering a streamlined and easily implementable solution for large-scale models without detailed microphysics schemes, which would not otherwise support the explicit treatment of SIP. Section 2.2.4 is revised accordingly.

Lines 145--148: These tunings may require some more detail and clarification. For example, I have no idea what RPRCON is. And second, the last sentence comes out of nowhere. Third, I have no idea what you meant by cloud forcing --- you mean cloud radiative effects like you refer to them later or something different? Altogether, the section left me wondering: What did the other preceding changes do to the forcing? Is the ~1 W/m2 solely from these three tunings or from everything together?

As mentioned in the Section 2.2.6, tuning is done to the model after major updates to minimize the radiative imbalance at the TOA and at the surface (global mean net radiative flux at the TOA, TOA and surface SW and LW flux, cloud radiative effects in LW and SW) in

AMIP simulations with CERES-EBAF data set as reference. There is a set of tuning parameters associated with each model and in this study, only three of these parameters were adjusted. They were 1. RPRCON, which is, as mentioned in the paragraph, the conversion efficiency, i.e. the rate of autoconversion of cloud water to rain, 2. the critical radius for autoconversion and 3. the standard deviation of the updraft velocity distribution in the cloud droplet activation.

The combined model updates led to a TOA net radiation imbalance of more than -3 W/m2 with the old settings of the tuning parameters. In contrast, the CMIP6 version of the model had an imbalance of -0.46 W/m2. By adjusting the above parameters we were able to bring the model closer to a radiative balance and a net TOA imbalance of -0.7 W/m2.

Section 2.2.6 is revised to incorporate the above mentioned details.

Line 181: What is meant by rotated here?
The different modes in MCA are obtained through singular value decomposition (SVD) of the covariance matrix between the two input fields. These modes are orthogonal and explain the maximum covariance between the fields. But, the resulting patterns can be complex and difficult to interpret physically. However, in the 'Rotated' MCA, the orthogonality constraint is relaxed to obtain easier to interpret patterns in terms of physical processes. In this case, the patterns are simplified by maximizing the variance across each mode. In our study, Varimax rotation technique was used. This has been clarified in the revised text.

Line 189: What is meant by an invariant quantity here?
This means that, in this case, the squared covariance fraction remains constant even if the modes are modified or transformed in some way. This has been added to the revised text.

Lines 190--214: I do not necessarily see this as needed in the main manuscript, but if the authors deem it necessary, I would advise to reorganize the section. I would start with the raw data X, Y (as an aside, what do a x n and b x n actually mean?) and then develop Ax and Ay using the SVD before defining the correlations. I would also add --- in equation form --- what "Corr" means. Finally, I would not have hom, het, Corr, and SCF be italicized.
This section has been rewritten based on the reviewer's suggestions.

Lines 216--220: Is this AMIP run completely compliant with published protocols? If not, state deviations. It would be good to simply state the full description of forcers and settings. Also, 1980--2020 AMIP vs 1980--2018 CMIP6?
Both AMIP simulations are fully compliant with the CMIP6 AMIP protocol, but have been extended beyond the CMIP6 historical period using anthropogenic and biomass burning emissions of ozone and aerosol precursors, GHG concentrations and other forcings following the CMIP6 protocol for the SSP2-4.5 scenario experiment. SSTs and sea ice fractions are prescribed using the CMIP6 version (1.1.8) of the PCMDI forcing data set, which extends out through 2021. We have clarified this in the text.

Figure 1 and associated text: In the bottom panel in the figure, please consider having the color bar go from -0.25 to +0.25 (i.e., symmetric). Also, any guess which processes (line 240) have led to the changes you're describing?

We understand the thinking behind having a symmetric colorbar, but we have left the figure as it is as a symmetric colorbar will not highlight the differences that are less than 10% as shown in the figure below:
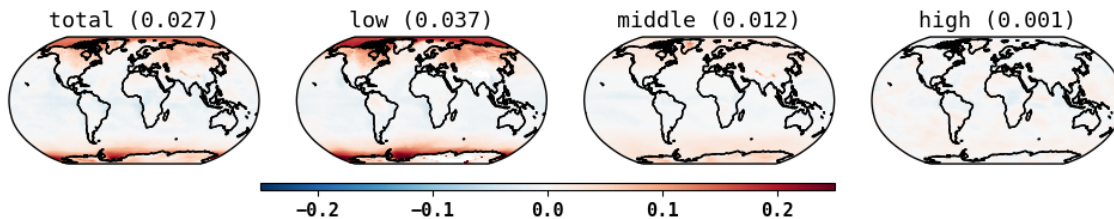
We have additionally added a sub plot showing the differences in the simulated cloud fraction with respect to CALIPSO observations as suggested by the second reviewer. With respect to ECE3-CMIP6 simulations, an increase in cloudiness is simulated in the high latitude regions. This increase mainly stems from the increase in low level cloud fraction. And this can be attributed to the model updates (for eg. the cloud droplet activation scheme) that targets liquid water clouds. This is mentioned in the revised manuscript.

Figure 2 and associated text: Consider having the same y-axis for all (maybe 0.0 to 1.0) and consider adding axis labels.

Following the review suggestion, the range of the y-axis is kept the same for all the subplots and the axis labels have been added to the figure. This is updated in the revised manuscript.

Figure 3 and associated text: Maybe I am more optimistic than you, the model seems to be doing okay for stratocumulus regions. Maybe show the spatial distribution of the difference if you want to highlight the difference?

Thank you for the suggestion. The difference plot and a description are added to the revised manuscript. The difference plot captures more than what meets the eye and it is a good idea to show the difference plots.

Here, we can actually see that the total cloud fraction is underestimated by the model over the tropics and the mid latitudes by 20-25%, however, the water cloud fraction is slightly overestimated by 10% and the model simulates well the ice cloud fraction over these regions compared to the MODIS observations.

Figure 4 and associated text: I would have appreciated seeing the total means on these figures like you showed earlier. Again, consider having the difference plot with a symmetric color bar (-70 to +70 or so). Given the plots as well as the discussion, maybe it is worth showing the combined (SW + LW) CRE?

The global means are shown in the revised figure. Here, again a symmetric color bar will not highlight the negative CRE differences as clearly as it is presented now. The combined (SW+LW) plot would show a spatial pattern more similar to the SW-only plot and also, to reduce the number of subplots, this will not be included in the revised manuscript. We have used a different colorbar so that the white regions correspond to 'no changes'.

(b) Differences (ECE3-CMIP6 - CERES-EBAF obs.)

SW (2.07)   LW (-0.96)

(c) Differences (ECE3-FORCeS - CERES-EBAF obs.)
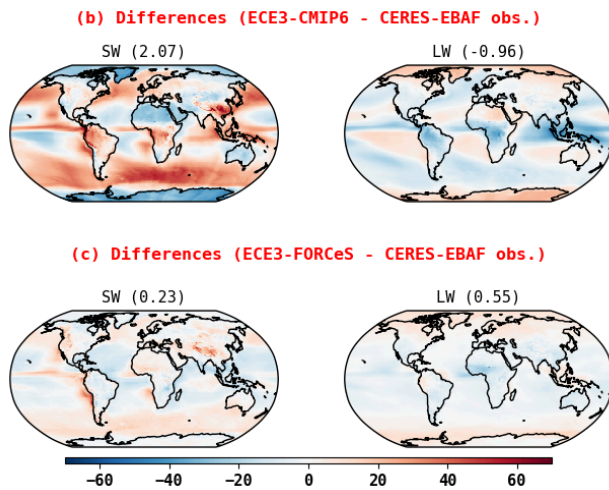
SW (0.23)   LW (0.55)

Figure 5 and associated text: As you say, these plots should be remade with better color bars. Also, because of the significant differences (underestimation by half), it may be worthwhile to give more reasoning and context. How worried should you be about this weakness of the model? Are there any tunings or strategies to remedy the situation? Have other modeling centers faced similar problems?

The figures have been revised with one colorbar for both CLWP and CIWP in the subplots. The updated model version has reduced the CLWP bias but increased the CIWP bias. Better model tuning could potentially improve cloud water path, both liquid and ice. We essentially re-tuned the model using a subset of the parameters identified in the tuning strategy of the CMIP6 version of the model, and we have clarified the tuning procedure in the main text. However, finding a new set of tuning parameters to improve clouds while maintaining radiation balance, cloud forcing, surface temperatures, precipitation patterns, etc., is challenging. A complete re-tuning was beyond the scope of this project and paper.. Future model developments aim to reduce biases through new parametrizations for updraft velocity and secondary ice production (RaFSIP v2). The extent to which these changes, along with re-tuning, could mitigate the biases requires further investigation. Model improvement is an ongoing process, and more effort is needed to further reduce biases.

Figure 6 and associated text: What's going on with MODIS near the poles? If no values are available there, maybe it is worth denoting that with white (like in Figure 7) as opposed to 0. Also, consider extending the color bar to 0.

The colorbar is extended to zero and the regions where MODIS observations are lacking are set to white. The minimum value of cloud droplet radius is 4 microns in the models.

Figure 7 and associated text: Oh, so this comparison kind of makes no sense. Why not add cloud-top diagnostics to your model? The MODIS simulator can do that or a simple cloud-top algorithm can do the job too. I would remove this comparison if there is no better comparison. There is no need to muddy the water with badly constructed comparison here in my opinion.

We agree with the reviewer here. We have removed this comparison as we do not have the corresponding COSP output for fairer CDNC evaluation.

***Figure 9 and associated text: I think it would be better to redo the figure and* ensure the color bars are the same for all AOD panels on the one hand, and all cloud fraction panels (ideally, 0 to 1) on the other hand.**
The figures are replotted incorporating the reviewers suggestions.

Line 393 and thereabouts: You're essentially describing the aerosol effective radiative forcing, but somehow only discussing two of the well known mechanisms. An example of a missing one is precipitation suppression. How did you exclude it from your two hypothesized processes? I am glad you're citing the Gryspeerdt et al. 2019 paper, but I am not sure your text gives a proper summary of the convoluted nature of all of this...
The precipitation suppression is indeed an important process. Here, we were thinking mainly from the top of the atmosphere perspective of cloud radiative effects. The changes in cloud microphysics and cloud cover have the largest impact on the changes in TOA CREs. Precipitation suppression, although not directly affecting TOA CREs, does impact the cloud lifetime and cloud cover. However, here we see that in the composite where PCs of AOD > 0 the absolute AOD is lower than the other AOD. We therefore suspect that the increase in cloud fraction is probably not linked to the increase in precipitation suppression.

Figure 14: Is the figure spreading on two pages intentionally? Maybe reorganize it so that it fits on one page?
The figures are intentionally spread on to two pages to ensure clarity.

The entirety of section 4: Please say more about the data sampling and processing. How many years were included? When the data was put in the statistical analysis framework, what was the frequency of sampling or was it just monthly/yearly means? What are the implications? From Figure 13, it appears you're using 40 years worth of data, monthly averaged? Maybe this was somewhere in the text but I missed?
Please note that the Section 4 has been reorganized as follows:
4 Results-II: Maximum Covariance Analysis: Covariability between AOD and CD effective radius
4.1 Regions selected for the analysis
4.2 MC analysis

Maximum covariance analysis is applied to AOD and CD radii to assess the covariability between these two variables. Since the analysis focuses on the months when the frequency of aerosol and cloud overlap is higher in the outflow regions, as explained in Section 4.1, the temporal sampling differs regionally. For instance, as shown in Table-1, we use all 492 months (41 years) of data for Eastern China. However, over the biomass burning regions of Africa and South America, the temporal sampling is limited to 122 and 246 months,

respectively. Since the focus of MCA analysis has been on testing the visibility of Twomey effect at a climate scale, the monthly means are analyzed. Before applying MCA, both the variables are deseasonalized and detrended based on the respective data for each region so as to ensure that this analysis captures the covariability between the two variables unbiased by the seasonal or long term trends.

The heterogeneous patterns, the patterns that bring out the areas of covariability will be presented for the three regions selected for the study. Only the regions exhibiting statistical significance at the 95% significance level are showcased. The following subsections detail the first three modes of the analysis over the regions selected for this study.

The entirety of section 4: Maybe putting lat/lon ticks (and their values) on the figures will be helpful?
Fig. 9 is replotted to clearly show the selected regions. Lat/lon ticks have been added to all the figures in section 4 in the revised manuscript.

The entirety of section 4: Maybe doing the same analysis on both model versions (the improved and the one before it) will be interesting?
We did actually plan to do this, but decided not to include such a comparative analysis in the end mainly for pragmatic reasons. First, we already did a thorough evaluation of cloud properties between those two versions of EC Earth. And second, we believe it will not add more value to the present analysis since the results suggest that AOD-CDradii co-variability and Cloud cover - CRE relationship will likely dominate in both model versions. In future, we do however plan to perform a comparative analysis with the satellite based observations.

Line 434 and thereabouts: Would you mind elaborating regarding what you mean here? For example, I think studying things at cloud top is likely better than studying them elsewhere; at the end of the day, that's where the radiation is reflected, no?
Yes, indeed the changes at the cloud top are important here from the top of the atmosphere perspective. We simply wanted to remind the readers that the typical limitations associated with the analysis of satellite based retrievals from the passive imagers do not apply in our case and that the co-variability signal we see is indeed physical.

(contd) Another, how do your model runs avoid the sampling issues related to simultaneous aerosol and cloud retrievals? How are you analyzing/sampling your model data?
In the model evaluation section, for comparisons with satellite data, we used COSP-derived variables whenever those were available so as to avoid issues related to sensitivity of satellite based instruments and time sampling. Furthermore, it is to be noted that, unlike the passive

satellite imagers that provide cloud information at the top and the total column aerosol measure such as AOD, the model estimates are available for each atmospheric layer.

Lines 437--438: Say more. Why could it be a good proxy? What makes you say so?

AOD has been used as a good proxy for the following reasons:

1. The Eastern China region selected for the co-variability study is the region where the pollution outflow just off the eastern coast is predominantly concentrated in the lower tropospheric levels.
2. Not all aerosols can serve as cloud condensation nuclei; for example, over the outflow regions over Eastern China, the aerosols are of more anthropogenic in origin and can be more efficient CCNs compared to the biomass burning aerosols over the other two regions.

Also, CCN was not one of the standard variables of EC-Earth3 output.

Lines 449: Based on what specifically can it be attributed? What you showed in the maximum covariance analysis? I am not sure if that's as clear cut as you're saying here, so I encourage further discussion.

This is explained in Section 4.5 in the original manuscript. This is supposed to be a separate paragraph and we have rephrased this sentence and summarized a bit more details as given below.

'We also further analyzed the implications for TOA SW CREs associated with this covariability. This was evaluated based on the PCs of AOD.'

Code and data: Thank you for making your analysis scripts readily available! Maybe mention that you're also including the MC code out there?

A sentence is added to the revised manuscript mentioning this.