

Author response to Referee 1

We kindly thank the referees for their relevant comments, suggestions, and corrections.

The responses are organized as follows: the reviewer's comment is in blue, our answers are in black, and the changes proposed for the revised manuscript are in italics (black for modified sentences, grey for unchanged sentences that have been copy-pasted here to remind the context). The numbering of the pages and lines corresponds to the preprint, not to the modified document. Last, we answer "Done" to all the comments suggesting a modification that we completely agree with, and that do not require any clarification in this document.

Before addressing the review itself, we have to highlight the update of the results since the preprint submission, as we made a small error on the radiative forcing calculation, more precisely on methane feedback: we computed the latter accounting for the decrease of methane lifetime due to reaction with OH, but we forgot to include the stratospheric loss and soil deposition.

In the abstract, the previous statement:

"The NO_x net ERF is systematically positive and ranges from 7.3 to 22.1 $mW m^{-2}$ among the different models (14.1–22.1 $mW m^{-2}$ without the least sensitive model)."

has been changed into:

"The NO_x net ERF is systematically positive with a model mean of 18.3 $mW m^{-2}$, ranging from 9.4 to 24.5 $mW m^{-2}$ among the different models. This net NO_x forcing is reduced by 35% and 43% accounting for the negative forcing arising from the formation of nitrate and sulfate particles, respectively."

Radiative forcings related to NO_x in Table S6 have been corrected as follows (the short-term NO_x effect on ozone has also been updated, but it is smaller than 0.2 $mW m^{-2}$):

ERF ($mW m^{-2}$)	EMAC- NO_x	LMDZ-INCA	MOZART3	OsloCTM3	GEOS-Chem
Former results					
Short-term ozone	27.7	43.0 [40.6]	42.0	34.0	56.0
CH ₄ direct effect	-13.2	-17.1	-12.9	-12.9	-24.8
Long-term ozone	-5.37	-6.96	-5.24	-5.24	-10.1

Stratospheric H ₂ O	-1.74	-2.25	-1.70	-1.70	-3.26
Total CH ₄	-20.4	-26.4	-19.9	-19.9	-38.2
Total	7.37	16.7	22.2	14.1	17.9
New results					
Short-term ozone	27.6	43.0 [40.6]	42.2	33.8	56.0
CH ₄ direct effect	-11.9	-15.4	-11.5	-11.6	-22.0
Long-term ozone	-4.83	-6.24	-4.68	-4.73	-8.93
Stratospheric H ₂ O	-1.56	-2.02	-1.51	-1.53	-2.89
Total CH ₄	-18.3	-23.6	-17.7	-17.9	-33.8
Total	9.36	19.4	24.5	15.9	22.2

We also added further information on the calculation of radiative forcings, in Section 2.3:

“To derive the radiative impact of aviation-induced atmospheric composition changes, we use concentration-based kernels to calculate the stratospherically adjusted ozone RF and the instantaneous top-of-the-atmosphere RF due to aerosol–radiation interactions (Skeie et al., 2020; Samset et al., 2011). To perform the RF calculations, ozone and aerosol data from all models were interpolated to the kernel resolution (2.25° x 2.25° and 60 vertical levels). For the calculation of the ozone column for each model, the air mass from OsloCTM3 was used, following the method in Skeie et al. (2020). The ozone RF calculations from the kernel have been found to compare favorably against offline radiative transfer model calculations in LMDZ-INCA and OsloCTM3. More generally, kernel-based estimates of ozone RF have been found to agree with those from full radiative transfer in previous applications (Lund et al., 2021). The change in methane volume mixing ratio and the associated radiative forcing are calculated based on the modelled change in methane total lifetime for each simulation. It assumes, in addition to the methane oxidation by OH, a stratospheric sink characterized by a 120-year lifetime and a soil sink characterized by a 160-year lifetime. This calculation is combined with the methane feedback factor (referring to the CH₄ feedback on its own lifetime) and an emission non-steady state factor, method described in Berntsen et al. (2005), Hodnebrog et al. (2012) and Terrenoire et al. (2022). The methane reference mixing ratio is fixed to 1834 ppb. The methane feedback factor ($f=1.45$) is taken as the model mean from a recent model intercomparison (Sand et al., 2023). We use a non-steady state factor to correct for the fact that due to its long lifetime, methane steady state is not reached, such as assuming steady-state to derive the radiative forcing overstates the response (Grewe and Stenke, 2008). This non-steady factor was recently recalculated by Bellouin et al. (in prep.) to be 0.680 for present-day conditions. From this methane mixing ratio change, the methane RF is calculated

using the simplified equation from Etminan et al. (2016). The indirect long-term ozone and stratospheric water vapour RFs are calculated based on the change in methane mixing ratio adopting the normalized forcings from a recent model intercomparison (Sand et al., 2023). For long-term ozone we use a normalized forcing of $0.180 \text{ W m}^{-2} \text{ ppbCH}_4^{-1}$ and, for stratospheric water, a normalized forcing of $0.058 \text{ W m}^{-2} \text{ ppbCH}_4^{-1}$.

The calculated ozone and methane RFs are converted to ERFs based on the efficacies provided by Lee et al. (2021) (1.370 for the short-lived ozone forcing and 1.180 for the methane direct and indirect forcings). For aerosols, the kernel includes rapid adjustments for BC, thus representing the ERF, while the ERF/RF ratio is assumed to be 1.0 for the scattering aerosols due to lack of other information (Lee et al., 2021).”

Another clarification has been added at the end of the paragraph on aerosol RFs commenting Fig. 11:

“[...] The RFs shown in Fig. 11 and discussed in this section are converted to ERFs to enable comparison with other studies in the following paragraphs (see Sect. 2.3). It is worth noting that in our modelling set-up, the nitrate particles forcing is not only related to aircraft NO_x emissions driving the formation of ammonium-nitrate particles (NH_4NO_3). It is also affected by the oxidation of sulfur dioxide (SO_2) into sulfate particles due to NO_x -induced OH formation. This additional sulfate then collides with ammonia to form ammonium-sulfate ($(\text{NH}_4)_2\text{SO}_4$) particles, thus competing with nitric acid to react with ammonia and thus to form nitrate aerosol. Terrenoire et al. (2022) show that the nitrate forcing is more negative (or less positive) when one accounts for the aircraft NO_x effects on aerosols. With this limitation in mind, we note that the nitrate forcing would reduce the net NO_x forcing from a model mean forcing of 18.3 mW m^{-2} to 11.9 mW m^{-2} .”

And in the conclusion:

“The net aerosol ERF_{ari} varies between -6.5 and -17.8 mW m^{-2} (multi-model mean of -11.6 mW m^{-2}), with the formation of sulfate and nitrate particles that reduces the net NO_x forcing by 43 % and 35 %, respectively.”

Concerning the introduction, we also point out the division of the second (and long) paragraph into two smaller paragraphs: one more general for the non- CO_2 emissions, and the second one more specific on the consequences of aviation NO_x .

Last, in the end, six figures have been added into the Supplement, both from the reviewers’ suggestions (Figs. S6, S9, S11–S13), and from our own initiative, as we found it would be useful for the reader (Fig. S7).

General comments

- This paper provides an important addition to the existing literature concerning the modelled effects of aviation NO_x pollution, which has high practical and policy relevance given the likely evolution of air transport.

- This paper does not itself add the strength of validating model predictions against observations. However, the referenced (published) companion paper (Cohen, 2025) does address how 4 of the 5 models compare against IAGOS observations. GEOS-Chem is the exception there though, which is often an interesting outlier in this paper.

- The overall climate forcing effect of aviation NO_x is still not well-constrained by this study, as the authors discuss, but it usefully adds results and analysis from further models. The authors should make clear what is new in this study in terms of models, date ranges and emissions covered, etc to distinguish it from previous work.

We totally agree with this important suggestion, made by both reviewers.

As written in the response to Referee 2, we added some additional remarks, as follows:

- In the introduction, in the third (formerly second) paragraph, we clarified two points addressed by our study, about the estimates from Lee et al. (2021). One concerns the formation of aerosols due to aircraft NO_x, and the other one concerns the heterogeneity of the studies reviewed by Lee et al. (2021), which is a strategy different than ours as it synthesizes 60 estimates taken from 20 studies.

“[...] Since methane is an ozone precursor in the troposphere and a water vapor precursor in the stratosphere, its increased sink in the short term decreases the production of these two species during the decade following the emission, thus increasing the cooling term due to methane destruction (e.g. Myhre et al., 2011). Through the production of tropospheric OH, aircraft NO_x also promotes the formation of sulfate and nitrate (NO₃) particles (because of an enhanced oxidation of SO₂ and NO_x), thus acting as an additional cooling factor (Brasseur et al., 2016; Terrenoire et al., 2022). Based on the existing literature, Lee et al. (2021) estimated the net aviation NO_x impact to be an effective radiative forcing (ERF) of 17.5 [0.6–28.5] mW m⁻² for the year 2018, resulting from a positive ERF from short-term ozone of 49.3 [33–76] mW m⁻² and a negative ERF from long-term methane decrease of -34.9 [-65 – -25] mW m⁻², thus highlighting high uncertainties for both processes. However, Lee et al. (2021) do not account for the NO_x effect on aerosol formation, as they point out the limited number of studies on this topic and the large associated uncertainties. Also, the assessment by Lee et al. (2021) is based on an important set of modelling studies which have been harmonised regarding time periods and aircraft emissions in order to account solely for inter-model differences in both chemistry and radiation. In this study we revisit the aircraft NO_x perturbations on chemistry and climate based on the latest generation of five global chemistry-climate/transport models and on a common modelling protocol regarding surface and aircraft emissions and time period covered. This further reduces the need for harmonisation of the model results as a post-treatment and allows to focus on the inter-model differences in the treatment of chemistry and dynamics.”

- In the introduction, at the end of the fourth paragraph, formerly the third, which reviews the literature:

“[...] processes not accounted for, as in Brasseur et al. (2016). The latter study also shows the impact of aircraft on gaseous and aerosol species, but not necessarily with the same models, thus limiting the interpretation of the results. Last, the aviation impact on nitrate aerosol is relatively new in the literature, and most of all, there is no model intercomparison regarding its perturbation due to aircraft emissions.”

- In the introduction, at the end of the 5th (formerly 4th) paragraph, we chose a more affirmative formulation regarding our explanation for the spatial pattern of the ozone perturbation:

“The objectives of this study are (1) to provide an overview of the methodology of the harmonized multi-model study, (2) to present aviation-induced changes in the concentration of reactive species and the extent to which models agree, as well as specific differences between individual models, including an evaluation of the linearity of aviation induced effects. Thanks to the ancillary variables provided by two of the models, our intercomparison is the first to suggest an explanation for the pattern of ozone changes in response to aircraft NO_x. Finally, (3) to provide estimates of the radiative effects of the simulated aviation-induced ozone, methane, and aerosol changes.”

This sentence replaces the former one:

“Based on ancillary variables provided by some of the models, we suggest additional explanations for the model results.”

- In the summary and conclusions section, first paragraph:

“In light of increasing air traffic, we perform and document a new multi-model assessment of the atmospheric composition response to aviation NO_x and aerosol/aerosol precursors emissions, and the associated radiative forcing of climate, notably providing the first model intercomparison on the impact of aviation on nitrate aerosol. The purpose is to refine estimates of the impact of aviation by limiting differences between the models in their implementations, for a better understanding of the intermodel variability in the results. We present a model intercomparison involving five state-of-the-art chemistry-transport models (CTM) or chemistry-climate models (CCM). For this study, each participating model provides a set of present-day runs, including at least one reference run with all the anthropogenic emissions, and one perturbation run without aviation emissions, using the same anthropogenic and biomass-burning emission inventories. The models used to estimate the impact of aviation-induced aerosols are also applied to the gaseous phase, thus maximizing the consistency between the estimates for gaseous and aerosol species. For all models, the effective radiative forcings are calculated with the same radiative code, the same grid, and the same feedback factors.”

- Same section, second paragraph:

“With the same common protocol applied to the models, our net NO_x ERF estimate and range, as well as the relative distribution between individual contributions from ozone, methane, and water vapor changes to the net NO_x ERF, remain similar to previous studies (normalized to present-day emissions).”

- There is probably less existing literature on the climate forcing from aerosols resulting from aviation pollution, so that novelty should be highlighted.

We agree with Referee 1. Additionally to Referee 2’s suggestions on the abstract regarding aerosols too, we added more context:

“This work shows encouraging results regarding our confidence in aviation NO_x-induced ozone response because of a good model agreement. To a lesser extent, some similarities in the results regarding aerosols are also encouraging, given the few existing model intercomparisons on this topic. However, results also highlight [...]”

Specific comments

- Abstract L28: "because of a better model agreement." better than what, previous comparisons? Or just "good model agreement"?

“Good model agreement” is a better formulation, we changed it in the abstract. Thanks for the suggestion.

- Section 2.1 L104: "the pairs of mixing states correspond to the hydrophilic-hydrophobic dichotomy" I understand that organic particles may start hydrophobic but become hydrophilic, but is this "dichotomy" a well-known term here? I'm not an aerosol expert but don't recognise it. Others may benefit from more explanation here too.

No, the term “dichotomy” is not a technical term. We clarified as follows:

“For each model described in Table 2, the pairs of mixing states correspond to the hydrophilic or hydrophobic state (with hydrophobic particles that can evolve into hydrophilic through ageing processes), and the three mixing states for EMAC-aer also include a mixed-particles category. It applies to both black carbon and organic carbon.”

- Introduction L84: "the recent inventories used by CMIP6 for anthropogenic surface and aircraft emissions", but CMIP6 data only runs to the end of 2014 (or 2015?); the referenced McDuffie 2020 work extends CMIP6 for a few years as CEDS(GBD-MAPS). See also L255.

We clarified, as for the next comment.

- Section 2.2 L255: "The historical anthropogenic emissions are taken from the Community Emissions Data System inventory CEDSv2 (McDuffie et al., 2020; O’Rourke et al., 2021,

regarding NO_x, SO₂, and BC emissions)." Having a look at McDuffie 2020 they only go to 2017, while the O'Rourke 2021 reference is for CEDS v_2021_04_21 which seems to go out to 2019. As some of the models run to 2018 or 2019, does that mean that the emissions for some species (those not in your O'Rourke list) were extrapolated or otherwise estimated to cover 2018 and 2019? A brief explanation of that would be good if so.

Thanks to Referee 1 for pointing out this confusing inconsistency. To clarify: at the time we made the simulations, the latest available dataset was the CEDS emission files from Hoesly et al. (2018), ending in 2014 (included). Hence the use of the emissions taken from the future scenarios of Gidden et al. (2019). Concerning the three mentioned species (NO_x, SO₂ and BC), it was unclear: those are the three species taken into account for aircraft emissions.

We modified the paragraph as follows:

"The historical anthropogenic emissions are taken from the Community Emissions Data System inventory CEDS (Hoesly et al., 2018). Regarding aviation emissions (represented here by NO_x, SO₂, and BC), these files resolve aviation emissions with 25 vertical levels (or injection heights) from the surface up to 15 km."

- Section 2.2 L260: "For these years (2015–2018)," but from Table 2 the GEOS-Chem model ran for 2019, so should it be 2015–2019 here? Or maybe GEOS-Chem used different emissions here too? (Later: more on GEOS-Chem emissions being different in section 2.3 L304, maybe worth a cross-reference)

We added a mention of 2019 in GEOS-Chem:

"For these years (2015–2018), the differences between the scenarios remain small (less than 6 % for NO_x), as are the differences with the year 2014 in the CEDS inventory, given that the scenarios data sets have been harmonized with the historical data sets to ensure a consistent evolution before and after this transition year (further information for the year 2019 in GEOS-Chem is available below)."

- Section 2.2 L280: "We rescaled the EMAC-aer perturbations using the NO_x emissions from CEDS for the period 2014–2018 (and the same rescaling factor for every species)," I'm not concerned that you had to do something special there, but don't really understand how that rescaling worked. E.g. for SEN100 I would expect you just zeroed out the aviation emissions in use by this model. So what is the "rescaling" of the perturbations? (Later: maybe related to section 2.3 around L294? In which case could cross-reference that section.)

We did not notice that we were mentioning the rescaling process before its definition, which is redundant and confusing. We therefore removed this sentence, as it is explained in detail a few lines after.

- Section 3.1 L391: "Due to stratospheric intrusions, an extension of the mean ozone perturbation is visible at low latitudes, downward and equatorward." Is this really about

stratospheric intrusion? If we had a global intrusion of that shape and size, surely it would be a named phenomenon... maybe it is partly. But we can already faintly see enhanced NO_x in the same area in Figure 4. So maybe it is largely created there in situ. After all, many flight routes following approximate great circles will climb and descend from lower-latitude airports but then (for most of the flight) cruise at higher altitude and further north, giving the same kind of "diagonal" pattern on these latitude-altitude plots. (Later: at L479 you make such a case for aerosols, though with actual near-surface emissions there too.)

We agree with Referee 1 on this comment. We changed our statement as follows:

“An extension of the mean ozone perturbation is visible at low latitudes (as for NO_x, to a lesser extent), downward and equatorward, which can be linked to the aircraft trajectories and to subsidence motions.”

- Section 3.1 L392: it might be interesting to comment on the somewhat different distribution of OH, which is most enhanced inside the UT compared to NO_x and O₃ in the LS. I guess because there is much more H₂O below the tropopause. (You come back to OH at L412+ but without venturing an explanation for its different distribution.) (Also return to OH around L444, could mention further discussion "below" at least here.)

The mention for further discussion below has been added as follows:

“In spring, the models generally exhibit a dipole structure with positive values centered near 40°N, mostly below the tropopause as well, and a negative response at high latitudes in the LMS (further discussion on chemical mechanisms responsible for the perturbation patterns is available below). The negative values are more pronounced with OsloCTM3, ...”

We also add a clarification that H₂O is not the only source of OH:

“The net OH response is higher with LMDZ-INCA (as expected from Fig. 6), characterized by a stronger positive response in the mid-latitudes and a weaker negative response in the high-latitude LMS. This positive OH response is consistent with the H₂O tropospheric background being the greatest in LMDZ-INCA, though the hydroperoxyl radical (HO₂) is another source of OH in the UT, which we did not investigate here. Last, [...]”

- Section 3.1 L409-410: reading the text and looking at Figure 5 side by side, two of the quoted numbers seem mismatched by a couple of tenths of a ppbv (6.7 vs 6.9, 13.0 vs 13.2). The differences may not be scientifically significant but they should agree, if only to make relating the text to the plots more straightforward, and anything >0.1 difference cannot just be a difference in rounding here. Perhaps the text was written against an older version of the plots or vice versa?

We thank Referee 1 for pointing out this inconsistency which is, indeed, probably due to a

last update made in the figures but not in the corresponding paragraph. All the values have been corrected now, and the statements adjusted, as follows:

“Some differences are well visible between the models. Contrasting with the ozone perturbation generally in the altitude range of 8–12 km, the OsloCTM3 ozone response is located lower in the troposphere, in the range ~5–10 km. In June, these lower altitudes in ozone perturbation are characterized by the strongest NO_x perturbation in the LMS, with a zonal mean above 120 ppt at all latitudes beyond 45° N. By contrast, the other models do not reach 100 ppt (LMDZ-INCA, GEOS-Chem), or very locally (EMAC-NO_x, MOZART3). In terms of mixing ratio, the maximum value of the ozone response ranges between 3.0 and 6.7 ppb in January, with three models relatively similar to each other (EMAC-NO_x, LMDZ-INCA, and MOZART3, within a range of 4.2–4.6 ppb). During the seasonal peak, the maximum value exhibits more discrepancies, with two models having lower maxima (4.1 and 6.9 ppb for OsloCTM3 and EMAC-NO_x respectively), two models having higher maxima (14.2 and 13.2 ppb for MOZART3 and GEOS-Chem), and one model having intermediate values (10.5 ppb for LMDZ-INCA).”

- Section 3.1 L420 "substantially weaker... -38 %" weaker than what, MOZART3?

We realise that it was not clear. It is explained now:

“With MOZART3, the maximum ozone response during JJA remains similar between the two studies, and 16 % lesser in DJF in the current study. With EMAC-NO_x and OsloCTM3 however, still in the current study compared to Søvde et al. (2014), the maximum ozone response is substantially weaker in both seasons (-38 % and -47 % in summer, and -39 % and -51 % in winter).”

- Section 3.1 L403-L422: this is a long paragraph that starts with O₃ (fig 5), switches to OH (fig 6), then backtracks to O₃ (fig 5). It might be a smoother experience for the reader if you moved the OH discussion to be a new, following paragraph.

Done.

- Section 3.1 is pretty long overall, and without any lower-level headings you can't cross-reference easily between parts of it. Maybe split it up into subsections e.g. for the different species discussed?

The species are regularly compared together, almost in each paragraph, which makes it hard to separate them into different subsections. Instead, we divided Section 3.1 into subsections linked to the kind of analysis we make:

3.1.1 Seasonal cycles in the UTLS (linked to Fig. 1: Hovmöller diagrams)

3.1.2 Geographical distribution in the UTLS (linked to Figs. 2 and 3: maps)

3.1.3 Zonal cross sections (linked to Figs. 4–6)

3.1.4 Involved chemical mechanisms

3.1.5 Dependence on the chemical background

- Section 3.1 L452 "stronger in models with lower NO_x" I agree, also lower NO_y more generally so could mention that too

We added an anecdotic comment noticing it, as follows:

"It appears that the responses in ozone and OH in the UTLS are generally stronger in models with lower NO_x (and lower NO_y) UTLS background, though this is not the only factor controlling O₃ and OH sensitivities to NO_x emissions."

- Section 3.1 L453 "The OH response increases with H₂O background" the GEOS-Chem point is a real outlier here; I see no correlation if I include that as an equally valid datapoint. You do mention in the next sentence that it is much higher but I'm left not really knowing the cause of that. If GEOS-Chem is excluded as a special case, then arguably you do have correlations with background CH₄ and O₃! So we need a good reason to include it, or not, when considering OH at least.

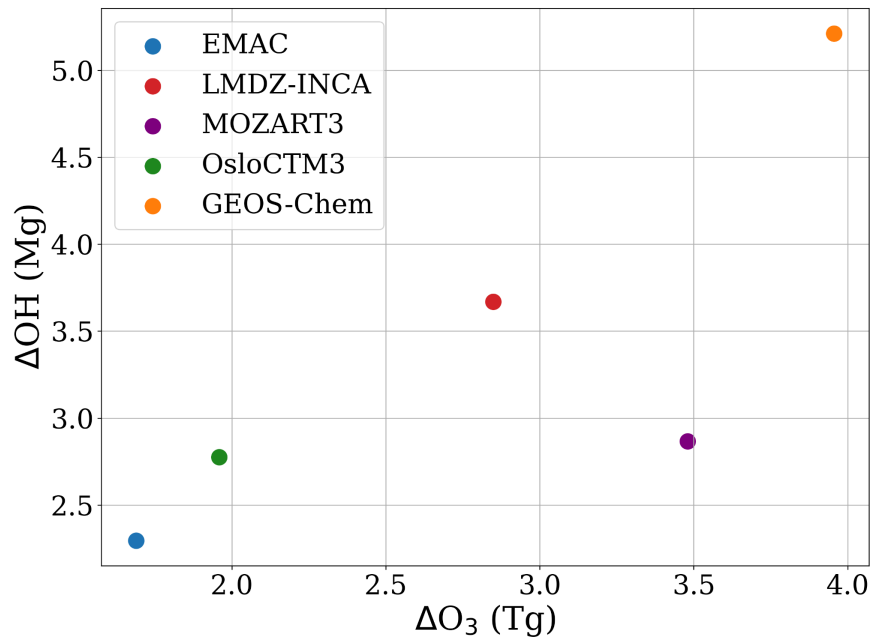
We added a note on the GEOS-Chem "exception", possibly due to substantial differences in the spatial distribution of NO_x emissions. Following Referee 2's request, we added maps showing partial columns for NO_x emissions, and it appears that GEOS-Chem has more balanced emissions between West and East hemispheres, whereas the other models have much more emissions in the Atlantic area (including North America, North Atlantic and western Europe) than in Asia (mainly East Asia, South Asia, and Siberia).

"The OH response is correlated with the ozone response, and with the H₂O background if not for GEOS-Chem. It is higher with GEOS-Chem notably because this model specifically does not show an OH negative response in the polar LMS. One possible explanation comes from the spatial distribution in NO_x emissions that differs between Fliht radar24 and CEDS inventories as mentioned in Section 2.2, leaving part of the emitted NO_x evolving in different chemical regimes between GEOS-Chem and the other models. On the opposite, the OH response is lower with EMAC-NO_x as both the ozone response and the H₂O background are relatively low. The OH response is comparable between MOZART3 and OsloCTM3, because of their strong negative response in the LMS, as seen in Fig. 6. The net OH response is higher with LMDZ-INCA (as expected from Fig. 6), characterized by a stronger positive response in the mid-latitudes and a weaker negative response in the high-latitude LMS. This positive OH response is consistent with the H₂O tropospheric background being the greatest in LMDZ-INCA."

- Section 3.1 L454 "OH response... is correlated with the ozone response." I completely agree, but we can't really see that from figure 7, because there is no plot of delta(OH) against

delta(O3). I plotted it myself with approximate values from figure 7 for each model and get a nice correlation (just MOZART3 a bit of an outlier from an otherwise straight line, as it happens, though I wouldn't draw any great significance from that). I suggest you add such a plot.

We agree with Referee 1 that this deserves a figure to make this statement more evident. We added the following figure in the Supplement, as the new Fig. S6.



- Section 3.2 L480: "and due to subsidence"; I'm no expert on this but perhaps that should be couched with some uncertainty, as you haven't really quantified whether there is more BC close to the ground because that's where it was emitted or because it sinks. (In principle maybe you could by a comparison with the input emissions distribution fed into the models.) Out of interest a quick web search gave a figure of 0.5 km/month as a subsidence rate in <https://doi.org/10.1126/science.aax1748> (but for the stratosphere?). (The BC lifetimes you later give around L492 might then suggest it disappears too quickly to subside much.)

Thanks to Referee 1 for this investigation. The paper cited here (Yu et al., 2019) apparently concerns stratospheric dynamics, which does comprise subsidence but with different characteristics than in the troposphere. But to be less affirmative, we added "possibly" before "due to subsidence".

- Section 3.2 L480: "The LMDZ-INCA and GEOS-Chem models show similar responses, with a maximum in April." Looks like it is higher for both in May not April, if my reading of the colours in Fig 8 is correct.

Though it is less visible with GEOS-Chem, the figure shows a higher maximum in April, at cruise levels at the mid-latitudes, with a small orange-red area.

- Section 3.2 L523: Terrenoire (2022) could get a mention too in this paragraph about previous aerosol perturbation work.

It is true, but Terrenoire et al. (2022) uses LMDZ-INCA with the same version as in our study, which would bias the comparison with the literature. We forgot to mention that this paragraph concerned the studies using a different model than the ones in our study, it is done now.

“Among the studies using different models than this paper, Unger et al. (2013) present the same spatial pattern [...]”

- Section 3.2 L535: "are more significant" maybe "are more significant by mass" say, as the BC might be more significant (say) in health terms?

Done.

- Section 4 L556/Figure 11: I presume the plots shown are a global average, but it would be good to state this explicitly. Also these ones are not month-specific, so are they annual averages? (I guess so from the "in most other months" comment about GEOS-Chem.)

The text has been updated to *‘Fig 11 shows the global annual average vertical profiles’*, as in Figure 11 caption.

- Section 4 L587: I just want to check I've understood the units correctly in "a range of 1.22–1.26 % (TgN yr⁻¹)-1 in methane lifetime". So for example if my no-aviation CH₄ lifetime was 10 years, it would decrease at the low end by $(1.22/100) \times 10 = 0.122$ years for each Tg of N aviation emissions? No need to clarify anything unless that interpretation is incorrect. Though should make it clear it is a decrease (could make numbers negative to emphasise that, as they are in Table 5). Also Table 5 has (-)1.25% not 1.26% for EMAC-NO_x, another rounding difference I guess.

Yes, that interpretation of the units is correct. We have updated to emphasise that it is a decrease: *“a range of -1.22 to -1.26 %”*. Lastly: we apologize for confusion, the right value is -1.26%.

- Section 4 L590: looks like an error in "GEOS-Chem sensitivity is ~190 % greater", should be ~90% ($= (2.36 - 1.23) / 1.23 \times 100$, taking 1.23 as the approx average of the lower 3 model values)

Done.

- Section 5 L659: "and a moderate perturbation (> 40 ppt) confined to the northern midlatitudes in winter" looking back at Fig 2 I'd say that's also true only for the high-traffic areas, which isn't quite clear here

Yes, it was implicit, as the 3 high-traffic areas are located in the northern mid-latitudes. We added “all of them in the mid-latitudes”, in order to make it clear.

- Section 5 L667: "mean net aviation NO_x ERF of 15 mW m⁻²" but at L592 it is "mean value of 16 mW m⁻²", i.e. 15 vs 16, again just small differences but they should be consistent, especially with this headline number

Once again, thanks to Referee 1 for pointing out this inconsistency. We homogenized it with a mean value of 15.6 mW m⁻², and a range of [7.3, 22.1] mW m⁻².

Edit: with the update of ERFs mentioned at the beginning of this document, the new phrase is finally: “Overall, we estimate a positive net aviation NO_x ERF in these model experiments, with a multi-model mean value of 18.3 mW m⁻² and a range from 9.4 to 24.5 mW m⁻².”

- Supplement S3 Figure S6 [now Figure S8]: as the emissions inputs are the same apart from GEOS-Chem (explained L274), I wonder why the MOZART3 ENO_x has a notably weaker northern latitude band, whereas the others have similar intensity in two latitude bands?

The MOZART3 model uses annual emissions with monthly weight factors to simulate a seasonal cycle, though this cycle shows less magnitude than the monthly emissions used in the other models. The NO_x levels in MOZART3 thus appear to be ‘flatter’ and more evenly spread over time compared to what is observed in other models, but the global totals and other NO_x figures do not show anything unusual about MOZART3 in this regard.

- Supplement S4 Table 3: I find the "Total CH₄" values are ~1.54x the "CH₄ direct effect" values. At L316 it says "the methane feedback factor from Sand et al. (2023)". But Sand et al have a model mean CH₄ feedback factor of 1.45 not 1.54. Best check this, and also clarify in the table legend what "Total CH₄" means.

We clarified “Total CH₄“ in the caption of Table 3 (now in Supplement S6):

“The row ‘Total CH₄’ corresponds to the sum of all ERFs related to CH₄, i.e. CH₄ direct effect, long-term ozone, and stratospheric H₂O. The row ‘Total’ corresponds to the sum of the two rows showing the short-term ozone ERF and ‘Total CH₄’.”

We have also clarified the meaning of the methane feedback factor (Section 2.3, completed with further information on the ERF calculations), as it relates to the changes in methane lifetime due to changes in methane itself, and not the effects on long-term ozone and stratospheric water vapour (which is another quantity).

- Supplement S4 Table 3: The "Total" row has your headline NO_x net ERF values. I note these values include "CH₄ direct effect" and not "Total CH₄", otherwise they would all be

lower (0-15 mW/m² instead of 7-22 mW/m²). Again at L316 it mentions using the methane feedback factor (which I think gives the "Total CH₄" numbers), but those feedback-adjusted values do not seem to be used in the headline total ERFs. For the uninitiated (including me), could you explain the use of feedback-adjusted (or not) CH₄ ERF I guess in the main text?

In Table S3, the row "Total CH₄" corresponds to the sum of methane direct effect and its two indirect effects: long-term ozone decrease, and stratospheric water vapour decrease. Same for the former question, it has been clarified in the table caption.

Technical corrections

Many of this are very small grammatical points.

- Abstract L19: "and the uncertainty on their Effective Radiative Forcing (ERF) of climate" should be "uncertainty in their..." (in not on), and might be better to say "the uncertainty in their climate Effective Radiative Forcing"

Done.

- Abstract L28: "However, results also highlight" -> "However, the results also..."

Done.

- Introduction L41: "black carbon (BC) - or soot - particles" to be super-picky those should be em dashes not hyphens

Done.

- Section 2.4 L104: "mixted" presumably "mixed"

Done.

- Section 2.1.1 L158: "used to EMAC-aer" should be "used in EMAC-aer"

Done.

- Section 2.1.3 L189: "organic halogen compound" presumably "compounds"

Done.

- Section 2.2 L243: "without any aviation emission" should be "emissions"

Done.

- Section 2.3 L288: "derive 5-year averages for each month" might be clearer to say "calendar month" here, as I assume you mean you're making an annual climatology, e.g. averaging the 5 different Januarys together, etc?

Done. And yes, we confirm Referee 1's interpretation.

- Section 2.3 L291: should be "factor of 5"

Done.

- Section 3.1 L328,L334,L337/Table 4: in the text we have "9.9 TgO₃" (three times) but the table has "10.0" for GEOS-Chem, so I think the same number has been rounded in different ways with the same number of decimal places. Easier to relate the two if they are presented identically.

Done. We totally agree, and updated the text according to the table.

- Figure 1: maybe just a preprint thing but in the current PDF the quality isn't too good, looks like it has been compressed coarsely as a JPEG or something. But up to the journal what they

expect really!

We thank Referee 1 for noting it and for leaving us the choice about it, which we forward to the journal.

- Figure 1: I realise all the other stacked plots show the different models vertically going down the page, but this one shows models going horizontally across the page. It's not wrong, and I know it takes a lot of effort to change figures, but if you needed to rework it anyway for other reasons you might consider if it would be nicer to be consistent and have models going down and species going across here too.

We agree it can be confusing, but we found this configuration practical to compare the seasonality of the different species together.

- Section 3.1 L390: redundant "for ozone" in "For ozone, the perturbation even peaks ... for ozone"

True. We removed the first iteration.

- Section 3.1 L430: "for the three models" maybe "for all three models"?

Done.

- Section 3.1 L432: should make clear the direction of the methane lifetime change depending on the NO_x background

This sentence has been clarified as follows:

"Table S2 also shows that the aviation-induced decrease in methane lifetime is enhanced by a factor of 5% [...]"

- Section 3.2 L480: this paragraph is pretty long, maybe start a new one for SO₄?

We agree with Referee 1 that the paragraph is quite long, but only two sentences are dedicated to SO₄ only, as for BC only. All the rest of the paragraph (both before and after) treats BC and SO₄ together, so cutting the paragraph in 2 would be less consistent. But it is as long as the next paragraph, which equilibrates the text.

- Section 3.2 L485: "but also in the whole free troposphere. The maximum shifts..." it isn't totally clear to me whether the second sentence here is talking specifically about EMAC-aer still, or all the models. From the plots I think it is just EMAC-aer. Could clarify that e.g. by doing "but also in the whole free troposphere; (in this model?) the maximum shifts..."

Done (we confirm that this concerns EMAC-aer only).

- Section 3.2 L490: "can be caused by a number of factors, as the BC lifetime" intended "such as the BC lifetime"?

Done.

- Section 4 L638: "Last, " should be "Lastly, "

Done.

- Supplement L33: "in bracket" -> "in brackets"

Done.

Author response to Referee 2

We kindly thank the referees for their relevant comments, suggestions, and corrections.

The responses are organized as follows: the reviewer's comment is in blue, our answers are in black, and the changes proposed for the revised manuscript are in italics (black for modified sentences, grey for unchanged sentences that have been copy-pasted here to remind the context). The numbering of the pages and lines corresponds to the preprint, not to the modified document. Last, we answer "Done" to all the comments suggesting a modification that we completely agree with, and that do not require any clarification in this document.

Before addressing the review itself, we have to highlight the update of the results since the preprint submission, as we made a small error on the radiative forcing calculation, more precisely on methane feedback: we computed the latter accounting for the decrease of methane lifetime due to reaction with OH, but we forgot to include the stratospheric loss and soil deposition.

In the abstract, the previous statement:

"The NO_x net ERF is systematically positive and ranges from 7.3 to 22.1 mW m⁻² among the different models (14.1–22.1 mW m⁻² without the least sensitive model)."

has been changed into:

"The NO_x net ERF is systematically positive with a model mean of 18.3 mW m⁻², ranging from 9.4 to 24.5 mW m⁻² among the different models. This net NO_x forcing is reduced by 35% and 43% accounting for the negative forcing arising from the formation of nitrate and sulfate particles, respectively."

Radiative forcings related to NO_x in Table S6 have been corrected as follows (the short-term NO_x effect on ozone has also been updated, but it is smaller than 0.2 mW m⁻²):

ERF (mW m ⁻²)	EMAC-NO _x	LMDZ-INCA	MOZART3	OsloCTM3	GEOS-Chem
Former results					
Short-term ozone	27.7	43.0 [40.6]	42.0	34.0	56.0
CH ₄ direct effect	-13.2	-17.1	-12.9	-12.9	-24.8
Long-term ozone	-5.37	-6.96	-5.24	-5.24	-10.1

Stratospheric H ₂ O	-1.74	-2.25	-1.70	-1.70	-3.26
Total CH ₄	-20.4	-26.4	-19.9	-19.9	-38.2
Total	7.37	16.7	22.2	14.1	17.9
New results					
Short-term ozone	27.6	43.0 [40.6]	42.2	33.8	56.0
CH ₄ direct effect	-11.9	-15.4	-11.5	-11.6	-22.0
Long-term ozone	-4.83	-6.24	-4.68	-4.73	-8.93
Stratospheric H ₂ O	-1.56	-2.02	-1.51	-1.53	-2.89
Total CH ₄	-18.3	-23.6	-17.7	-17.9	-33.8
Total	9.36	19.4	24.5	15.9	22.2

We also added further information on the calculation of radiative forcings, in Section 2.3:

“To derive the radiative impact of aviation-induced atmospheric composition changes, we use concentration-based kernels to calculate the stratospherically adjusted ozone RF and the instantaneous top-of-the-atmosphere RF due to aerosol–radiation interactions (Skeie et al., 2020; Samset et al., 2011). To perform the RF calculations, ozone and aerosol data from all models were interpolated to the kernel resolution (2.25° x 2.25° and 60 vertical levels). For the calculation of the ozone column for each model, the air mass from OsloCTM3 was used, following the method in Skeie et al. (2020). The ozone RF calculations from the kernel have been found to compare favorably against offline radiative transfer model calculations in LMDZ-INCA and OsloCTM3. More generally, kernel-based estimates of ozone RF have been found to agree with those from full radiative transfer in previous applications (Lund et al., 2021). The change in methane volume mixing ratio and the associated radiative forcing are calculated based on the modelled change in methane total lifetime for each simulation. It assumes, in addition to the methane oxidation by OH, a stratospheric sink characterized by a 120-year lifetime and a soil sink characterized by a 160-year lifetime. This calculation is combined with the methane feedback factor (referring to the CH₄ feedback on its own lifetime) and an emission non-steady state factor, method described in Berntsen et al. (2005), Hodnebrog et al. (2012) and Terrenoire et al. (2022). The methane reference mixing ratio is fixed to 1834 ppb. The methane feedback factor (f=1.45) is taken as the model mean from a recent model intercomparison (Sand et al., 2023). We use a non-steady state factor to correct for the fact that due to its long lifetime, methane steady state is not reached, such as assuming steady-state to derive the radiative forcing overstates the response (Grewe and Stenke, 2008). This non-steady factor was recently recalculated by Bellouin et al. (in prep.) to be 0.680 for present-day conditions. From this methane mixing ratio change, the methane RF is calculated

using the simplified equation from Etminan et al. (2016). The indirect long-term ozone and stratospheric water vapour RFs are calculated based on the change in methane mixing ratio adopting the normalized forcings from a recent model intercomparison (Sand et al., 2023). For long-term ozone we use a normalized forcing of $0.180 \text{ W m}^{-2} \text{ ppbCH}_4^{-1}$ and, for stratospheric water, a normalized forcing of $0.058 \text{ W m}^{-2} \text{ ppbCH}_4^{-1}$.

The calculated ozone and methane RFs are converted to ERFs based on the efficacies provided by Lee et al. (2021) (1.370 for the short-lived ozone forcing and 1.180 for the methane direct and indirect forcings). For aerosols, the kernel includes rapid adjustments for BC, thus representing the ERF, while the ERF/RF ratio is assumed to be 1.0 for the scattering aerosols due to lack of other information (Lee et al., 2021).”

Another clarification has been added at the end of the paragraph on aerosol RFs commenting Fig. 11:

“[...] The RFs shown in Fig. 11 and discussed in this section are converted to ERFs to enable comparison with other studies in the following paragraphs (see Sect. 2.3). It is worth noting that in our modelling set-up, the nitrate particles forcing is not only related to aircraft NO_x emissions driving the formation of ammonium-nitrate particles (NH_4NO_3). It is also affected by the oxidation of sulfur dioxide (SO_2) into sulfate particles due to NO_x -induced OH formation. This additional sulfate then collides with ammonia to form ammonium-sulfate ($(\text{NH}_4)_2\text{SO}_4$) particles, thus competing with nitric acid to react with ammonia and thus to form nitrate aerosol. Terrenoire et al. (2022) show that the nitrate forcing is more negative (or less positive) when one accounts for the aircraft NO_x effects on aerosols. With this limitation in mind, we note that the nitrate forcing would reduce the net NO_x forcing from a model mean forcing of 18.3 mW m^{-2} to 11.9 mW m^{-2} .”

And in the conclusion:

“The net aerosol ERF_{ari} varies between -6.5 and -17.8 mW m^{-2} (multi-model mean of -11.6 mW m^{-2}), with the formation of sulfate and nitrate particles that reduces the net NO_x forcing by 43 % and 35 %, respectively.”

Concerning the introduction, we also point out the division of the second (and long) paragraph into two smaller paragraphs: one more general for the non- CO_2 emissions, and the second one more specific on the consequences of aviation NO_x .

Last, in the end, six figures have been added into the Supplement, both from the reviewers' suggestions (Figs. S6, S9, S11–S13), and from our own initiative, as we found it would be useful for the reader (Fig. S7).

The manuscript presents a new model intercomparison to quantify the impact of NO_x and aerosols emissions from aircraft on both atmospheric composition and climate. The

atmospheric composition piece includes ozone, NO_x, OH and aerosols – black carbon, sulfate and nitrate aerosols. The study includes three global chemistry transport models – MOZART3, OsloCTM3, GEOS-Chem – and two global chemistry climate models – EMAC, LMDZ-INCA. The authors found that aircraft NO_x emissions are causing an increase of ozone photochemical production in the free troposphere throughout the year and a decrease of ozone chemical loss in the high-latitude lowermost stratosphere in spring/early summer. They also found that the net Effective Radiative Forcing (ERF) of NO_x due to aircraft emissions is positive and vary with models between 7.3 to 22.1 mW m⁻², when the aircraft aerosol direct ERF is negative and vary between -6.5 and -17.8 mW m⁻².

The authors put in perspective their findings with prior studies and discuss the potential reasons of the uncertainties. They suggest ways to improve future model output to reduce those uncertainties.

The study is in the scope of ACP. It is a thorough study. The manuscript is well written and well organized with clear figures and tables.

I would support publication after the authors addressed the minor comments I listed below.

General Comments:

The manuscript would benefit from having a description of the aviation emission inventories used in this study: CEDS aviation emissions and Flightradar24, which could include the altitude of injection. Could the authors add a map of the aircraft NO_x emissions?

We totally agree with this suggestion. We added the partial columns of seasonal NO_x emissions in Supplement, Fig. S9, summed up between 150 and 400 hPa.

The authors do compare their results with former studies but it is not always clear when the study brings new findings on the impact of aircraft emissions on atmospheric composition and climate. Does this study bring to the community more accurate numbers with less uncertainties or new numbers and how (e.g. the impact of nitrate aerosols, the intensity of the direct effect of aerosols from aviation)? Could the authors write a separate section on this point? The section could also discuss the choice of the time-period and its potential impact on the results.

We totally agree with this important suggestion, made by both reviewers.

As written in the response to Referee 1, we added some additional remarks, as follows:

- In the introduction, in the third (formerly second) paragraph, we clarified two points addressed by our study, about the estimates from Lee et al. (2021). One concerns the formation of aerosols due to aircraft NO_x, and the other one concerns the

heterogeneity of the studies reviewed by Lee et al. (2021), which is a strategy different than ours as it synthesizes 60 estimates taken from 20 studies.

“[...] Since methane is an ozone precursor in the troposphere and a water vapor precursor in the stratosphere, its increased sink in the short term decreases the production of these two species during the decade following the emission, thus increasing the cooling term due to methane destruction (e.g. Myhre et al., 2011). Through the production of tropospheric OH, aircraft NO_x also promotes the formation of sulfate and nitrate (NO₃) particles (because of an enhanced oxidation of SO₂ and NO_x), thus acting as an additional cooling factor (Brasseur et al., 2016; Terrenoire et al., 2022). Based on the existing literature, Lee et al. (2021) estimated the net aviation NO_x impact to be an effective radiative forcing (ERF) of 17.5 [0.6–28.5] mW m⁻² for the year 2018, resulting from a positive ERF from short-term ozone of 49.3 [33–76] mW m⁻² and a negative ERF from long-term methane decrease of -34.9 [-65 – -25] mW m⁻², thus highlighting high uncertainties for both processes. However, Lee et al. (2021) do not account for the NO_x effect on aerosol formation, as they point out the limited number of studies on this topic and the large associated uncertainties. Also, the assessment by Lee et al. (2021) is based on an important set of modelling studies which have been harmonised regarding time periods and aircraft emissions in order to account solely for inter-model differences in both chemistry and radiation. In this study we revisit the aircraft NO_x perturbations on chemistry and climate based on the latest generation of five global chemistry-climate/transport models and on a common modelling protocol regarding surface and aircraft emissions and time period covered. This further reduces the need for harmonisation of the model results as a post-treatment and allows to focus on the inter-model differences in the treatment of chemistry and dynamics.”

- In the introduction, at the end of the fourth paragraph, formerly the third, which reviews the literature:

“[...] processes not accounted for, as in Brasseur et al. (2016). The latter study also shows the impact of aircraft on gaseous and aerosol species, but not necessarily with the same models, thus limiting the interpretation of the results. Last, the aviation impact on nitrate aerosol is relatively new in the literature, and most of all, there is no intermodel comparison regarding its perturbation due to aircraft emissions.”

- In the introduction, at the end of the 5th (formerly 4th) paragraph, we chose a more affirmative formulation regarding our explanation for the spatial pattern of the ozone perturbation:

“The objectives of this study are (1) to provide an overview of the methodology of the harmonized multi-model study, (2) to present aviation-induced changes in the concentration of reactive species and the extent to which models agree, as well as specific differences between individual models, including an evaluation of the linearity of aviation induced effects. Thanks to the ancillary variables provided by two of the models, our intercomparison is the first to suggest an explanation for the pattern of ozone changes in response to aircraft

NO_x. Finally, (3) to provide estimates of the radiative effects of the simulated aviation-induced ozone, methane, and aerosol changes.”

This sentence replaces the former one:

“Based on ancillary variables provided by some of the models, we suggest additional explanations for the model results.”

- In the summary and conclusions section, first paragraph:

“In light of increasing air traffic, we perform and document a new multi-model assessment of the atmospheric composition response to aviation NO_x and aerosol/aerosol precursors emissions, and the associated radiative forcing of climate, notably providing the first intermodel comparison on the impact of aviation on nitrate aerosol. The purpose is to refine estimates of the impact of aviation by limiting differences between the models in their implementations, for a better understanding of the intermodel variability in the results. We present a model intercomparison involving five state-of-the-art chemistry-transport models (CTM) or chemistry-climate models (CCM). For this study, each participating model provides a set of present-day runs, including at least one reference run with all the anthropogenic emissions, and one perturbation run without aviation emissions, using the same anthropogenic and biomass-burning emission inventories. The models used to estimate the impact of aviation-induced aerosols are also involved for the gaseous phase, thus maximizing consistency between the estimates for gaseous and aerosol species. For all models, the effective radiative forcings are calculated with the same radiative code, the same grid, and the same feedback factors.”

- Same section, second paragraph:

“With the same common protocol we applied to the models, our net NO_x ERF estimate and range, as well as the relative distribution between individual contributions from ozone, methane, and water vapor changes to the net NO_x ERF, remain similar to previous studies (normalized to present-day emissions).”

Specific Comments:

L.257: “CEDS aviation emissions” - Could the authors comment on the aircraft emissions inventory? Does it include altitude of injection?

Yes, the mass density of emitted pollutants is distributed onto constant pressure levels in the inventory. Each model then interpolates it onto its own grid. The mention of the injection altitude has been added at the first mention of aviation emissions in CEDS:

“The historical anthropogenic emissions are taken from the Community Emissions Data System inventory CEDSv2 (McDuffie et al., 2020; O’Rourke et al., 2021, regarding NO_x, SO₂, and BC emissions), which resolves aviation emissions with 25 vertical levels (or injection heights) from the surface up to 15 km.”

Would it be possible to add in the Supplement, the equivalent of Figure S1 but with the aircraft NO_x emissions?

Maps (partial columns) and zonal cross sections of aviation emissions are now represented in Figs. S7 and S8, respectively. We added the reference to these figures to make it visible.

L.260: “small” - Could the authors give a quantity?

If we compare the total anthropogenic emissions from three scenarios (SSP1-2.6, SSP2-4.5, SSP3-7.0), we get the following NO_x emissions, in TgNO₂:

Year	SSP1	SSP2	SSP3
2015	153.7	154.3	157.3
2016	151.8	153.0	159.0
2017	150.0	151.8	160.8
2018	148.1	150.5	162.5

The values provided in the source (SSP database, link below) only concern the years 2015 and 2020, so the values for 2016–2018 listed above are derived from a linear interpolation.

The mean values, once converted into TgN/yr, are respectively 48.0, 48.5, and 50.9, thus with a maximum difference of 5.7 %.

Source: SSP database, <https://tntcat.iiasa.ac.at/SspDb/dsd?Action=htmlpage&page=50>

It has been clarified in the text:

“For these years (2015–2018), the differences between the scenarios remain small (less than 6 % for NO_x), as are the differences with [...]”

L.274: “Flightradar24” – As for CEDS aviation emissions. Could the authors describe it more – because it is a key component of the study: altitude of injection, latitude/longitude? A paragraph dedicated to aircraft emission inventories as input for the models used in the study would be key.

A more complete representation of the emissions is now available, in Figs. S6–S8 in Supplement. The comparison between CEDS and Flightradar24 shows substantial differences in Figs. S7 and S8, notably with less emissions over North America and North Atlantic and more emissions over Asia, as explained below:

“Among the five models included in this paper, GEOS-Chem data is from pre-existing runs made for a different publication (Quadros et al., 2025) and is thus less consistent with the protocol. For this model, the monthly-averaged aircraft emissions are calculated from a list of flights provided by Flightradar24, as described by Quadros et al. (2022). For each

combination of aircraft type, origin and destination in the database, 3-d gridded fuel burn is calculated using a time-in-mode approach for landing-and-takeoff operations (Stettler et al., 2011) and the Base of Aircraft Data (BADA) 3.15 aircraft performance model for climb, cruise and descent phases of flight (Mouillet, 2019). Great-circle trajectories between airports are used, with a lateral inefficiency factor based on Seymour et al. (2020) applied to adjust for additional fuel used in actual trajectories. Constant cruise flight levels are used, with aircraft type specific values determined from a subset of flights for which Flightradar24 provided altitude at the start of cruise, based on ADS-B data. Emissions are calculated alongside with fuel burn using the Boeing Fuel Flow Method 2 (Baughcum et al., 1996), the FOA4 method for non-volatile particulate matter (ICAO, 2020, last access: April 2020), and data from the International Civil Aviation Organization (ICAO) Engine Emissions Databank (last access: Jan. 2021).

The difference between Flightradar24 and CEDS emissions is well visible in the maps shown in Fig. S9 in Supplement, with less NO_x emitted in West hemisphere and more in East hemisphere with Flightradar24. In terms of altitude, Figure S10 shows a continuum in the vertical distribution in CEDS emissions (at 9–11 km) but two distinct peaks with Flightradar24, the most important at 10 – 11 km (as with CEDS) and a secondary peak at 8 – 9 km, lower than most emissions in CEDS. Also, the runs concern only the year 2019.”

L.301-306 (a): Does the rescaling methodology implies that the vertical regridding for each model has been made with LMDZ-INCA grids as the reference? Could you clarify?

We realize that our formulation using “regridding” is confusing. The perturbations in chemical tracers have been calculated on each model’s original grid. The radiative forcings however have been computed after a regridding onto OsloCTM3’s to apply the same kernel resolution.

When we used the word “regridding” just before, it was in a context upwind from the simulations: for each model, the emission files have to be interpolated onto the model’s grid.

We rephrased as follows:

“As the aviation emissions provided as output by the models can be slightly different due to vertical regridding to model native resolutions and/or due to different years, we apply a rescaling factor to each model result to ensure the same amount of aircraft emissions. For a given model M, [...]”

“The aviation emissions provided as output by the models can be slightly different for two reasons. First, the regridding of the emission files to the model native resolutions, which are all different. Second, the differences in the simulation years. For this purpose, we apply a rescaling factor to each model result to ensure the same amount of aircraft emissions. For a given model M, [...]”

Concerning the averaging kernels, we clarified as follows:

“To perform the RF calculations, ozone and aerosol data from all models were interpolated to the kernel resolution (2.25° x 2.25° and 60 vertical levels, i.e. that of OsloCTM3).”

L.301-306 (b): Have the authors estimated the impact of the linear rescaling on GEOS-Chem’s results?

Unfortunately not. For this purpose, we would need to perform a new run with GEOS-Chem for the right period (2014–2018), and then to compare. On the contrary, we use a pre-existing run to avoid performing a new one.

L.323-324: Could the authors be more explicit on the calculation of the impact by explaining the ratio: $\Delta S/EN_{Ox}$ with ΔS the perturbation. How this perturbation is calculated? How EN_{Ox} is calculated?

We clarified the formalism used for the calculations by adding new equations (2, 3a, 3b, and 4) at the beginning of Section 3.

L.325-327: Where is the reference figure or table for the percentages (e.g. global NO_x perturbation ranges of emitted NO_x in terms of nitrogen mass).

These numbers are directly taken from Table 4. Where we can see a ΔNO_x -to- EN_{Ox} ratio of $0.006 \text{ TgN}/(\text{TgN yr}^{-1})$, we say it is 0.6 % of the mass of emitted NO_x during the year. We added a few clarifications to make it explicit:

“As seen in Table 4, the global NO_x perturbation ranges between 0.60 % (EMAC- NO_x) and 0.85 % (MOZART3) of the yearly emitted NO_x , in terms of nitrogen mass.”

L.327-328: “0.51 DU” – shouldn’t the unit be DU/NEU? / “0.90 DU” – shouldn’t it be 0.92 DU/NEU (number reported in Table 4)? / “9.9 TgO₃” – shouldn’t it be 10 Tg/NEU (number reported in Table 4)?

Yes, thanks to Referee 2 for pointing out this inconsistency. It has been corrected.

L.329: Would the authors have a reference figure or table for mixing ratios?

We realize that the sentence L329 is confusing (as it is not the reference mixing ratio but the perturbation in mixing ratio that can reach higher values with EMAC- NO_x than OsloCTM3), so we propose the following precision:

“The values are similar between LMDZ-INCA and MOZART3. OsloCTM3 shows a higher mean sensitivity compared to EMAC- NO_x though the perturbation for EMAC- NO_x mixing ratio reaches higher values (as seen later, in Fig. 5), which is due to the lower altitude of the OsloCTM3 response (hence a greater mass perturbation) and to its wider vertical range.”

We did not want to mention Figure 5 as we tried to avoid the reader going forth and back in the manuscript, but in the end, it might not be possible to avoid it totally.

L.330: “lower altitude of the OsloCTM3” – Could the author explain this characteristic in the method section? / “wider vertical range” - Could the width of the vertical range be added in Table 1?

We are not sure to understand Referee 2 on this question. The “wider vertical range” is a description of the results shown in a figure, not a method. As for the lower altitude of the ozone response for OsloCTM3.

L.332 a): ”AEDT” – Could the authors spell it out?

We spelled it out in the manuscript. It means Aviation Environmental Design Tool, provided by the Federal Aviation Administration (FAA).

L332 b): "between 2.8 and 11.2 Tg/(TgN.yr-1) from both offline and online models” – How these models compare with the CTMs of the study (emission inventories, chemical and physical processes)?

The chemical processes are similar, except for the nitrate chemistry, which has been further developed over the last 10 years. Concerning the emission inventories, they are similar to CEDS too, for aviation.

According to Olsen et al. (2013), the present-day (2006) non-aviation emissions are taken from RCP (Representative Concentration Pathway) 4.5, whereas our study imposes a complete and self-consistent emission inventory (CEDS).

L.334 a): “9.9” – wouldn’t be 10 Tg/NEU instead, according to Table 4?

Yes, we corrected this inconsistency.

L334 b): “The ozone sensitivity is lower in the current study” – Did the authors mean lower than what was found in Olsen et al. (2013) and Brasseur et al. (2016)? How ozone sensitivity is calculated? Is it different from the ozone burden sensitivity?

Yes, we mean lower than in these two studies, but only if we compare comparable things (offline models). Here, we do refer to ozone burden sensitivity, which we clarify:

“Compared to the current study (5.6–10.0 Tg/NEU), the inter-model range is similar. The ozone burden sensitivity is lower in our results than in these two studies, [...]”

L.335: “aviation NO_x emission is 36% greater in 2014–2018 than in 2006” – Do the authors have a reference, a figure or table to show this number for aviation NO_x emission change?

Yes, the emissions during 2006 are shown in Table 2 in Brasseur et al. (2016): 0.812 TgN/yr. Compared to the ENOx value in LMDZ-INCA (1.119 TgN/yr), it represents an increase of ~38 % (not 36 %). We clarified in the text:

“[...] the aviation NOx emission is 38 % greater in 2014–2018 (1.119 NEU) than in 2006 (0.812 NEU, as shown in Table 2 from Brasseur et al., 2016), [...]”

L.336: “GEOS-Chem is characterized by the highest response” – May be worth repeating 11.2 Tg/NEU in parenthesis here, if I correctly understand.

We agree and have added it to the text.

L.342-343: “In order to capture as much of the response for each model as possible” – Could you clarify? Could you explain the vertical range characteristics of the model and give information in Table 1 that you can refer here.

We rephrased this sentence as follows to make it clearer.

Former version:

“In order to capture as much of the response for each model as possible, the vertical average is made between 150 and 350 hPa.”

Last version:

“As the models can have different altitudes in the maximum ozone response to aviation NO_x emissions (as seen later for OsloCTM3 in Fig. 5), the vertical average is made between 150 and 350 hPa to capture as much of the response for each model as possible.”

L.344: “NOx response” - It would be worth defining *NOx response*, which seems to be called *NOx change* in Figure 1. It may be worth adding a paragraph in the methodology? Worth explaining why it is a different unit than in Table 4?

“Response” is more specific than “change”. We use “response” specifically for aviation, in a way equivalent to “aviation-induced change”. We replaced most “change” occurrences by “response” in order to avoid confusion. Sometimes, we keep “aviation-induced change” as the meaning is explicit with this formulation, thus with a low risk of confusion.

As to the unit, a global budget is more conveniently expressed in burden (total mass), whereas local values are more conveniently expressed as concentrations or mixing ratios.

L.371-372: “As expected, Fig. 3 shows a more homogeneous ozone perturbation” – Do the authors mean *more homogeneous than NOx*? Could the authors explain why it is expected?

Yes, that is the exact meaning of this phrase. The text has been updated as follows:

“Figure 3 shows a more homogeneous ozone perturbation compared to NO_x, as expected for a secondary pollutant.”

L.391: “for ozone” – Typo, “for ozone” is written twice in this sentence.

Thanks to Referee 2, we corrected this typo.

L.419: “lesser” – Could the authors specify which study gives a 16% lesser ozone response than the current study?

We added “*in the current study*” at the end of the sentence.

L.420: “weaker” - Could the authors specify which study gives a 16% lesser ozone response?

We clarified: “*still in the current study compared to Sovde et al. (2014)*”.

L.421: “perturbations” – Could the authors be specific and change to “perturbations of ozone”?

Done.

L.424: “studies that focus on the past decade” – Could the authors remind the reader of those studies by adding the references?

We clarified with the bracket ‘*(notably Sovde et al., 2014; Brasseur et al., 2016)*’.

L.431-432: It would be worth reminding the reference to Table S2 here.

Done, as follows:

“As a consequence, Table S2 also shows that the perturbation in methane lifetime is affected by a factor of 5% [...]”

L.435: “by EMAC-NO_x and LMDZ-INCA” – Could the authors remind the readers why not by the other models?

The clarification has been introduced:

“The causes for the spatial distribution of the ozone response have been investigated using the chemical production and loss terms for ozone from the two models that provided them as diagnostic output, i.e. EMAC-NO_x and LMDZ-INCA.”

L.440: “tends to reduce the ozone vertical gradient” – Could the authors back up this statement with a figure in the supplement?

We do not think it is worth another figure. As there is more ozone in the LMS (above) than in the UT (below), increasing ozone production in the UT tends to reduce the difference between the two layers, thus the vertical gradient.

Instead of an additional figure, we clarified with an explanation:

“First, the enhanced photochemical production in the UT tends to reduce the ozone vertical gradient (as background ozone is less abundant in the UT than above) [...]”

L.441: “Second, the chemical loss term (Fig. S2, in Supplement) increases in the UT” – The ozone loss seems to be maximum in the mid-troposphere (~ 600 hPa, Figure S2). It would be worth noting this and explain why focusing on the UT.

We agree that it can be confusing that we do not mention the maximum in the ozone loss term. We rephrased as follows:

“Second, though the chemical loss term (Fig. S2, in Supplement) increases in the mid-troposphere as ozone increases and water vapour remains relatively abundant (thus involving the reaction $O(^1D) + H_2O \rightarrow 2 OH$), the chemical loss term decreases in the high-latitude LMS during spring-early summer.”

L.443-446: This finding is very interesting. Could the authors elaborate more on this competition between O₃ and NO₂ to react with OH? Any reference?

We agree that this point warrants further development, notably because previous studies have not investigated the reasons for this pattern in ozone (and OH) perturbations, as it was not the main purpose of these experiments.

We can cite the book “Aeronomy of the middle atmosphere” written by Guy Brasseur and Susan Solomon (at least the 2005 edition), Fig. 5.71, p. 416, as shown below. It presents zonal cross sections of the share of stratospheric O_x destruction between different chemical families acting through catalytic cycles, in June, from the NASA/Goddard 2-D model. It is organized such that at a fixed location, the sum of the numbers across the 5 panels equals 100 %. Regarding our study, the area of interest is the lowermost stratosphere in the northern high latitudes. At this location, the major contributor to catalytic ozone loss is the HO_x family (HO₂+OH), accounting for ~80 %.

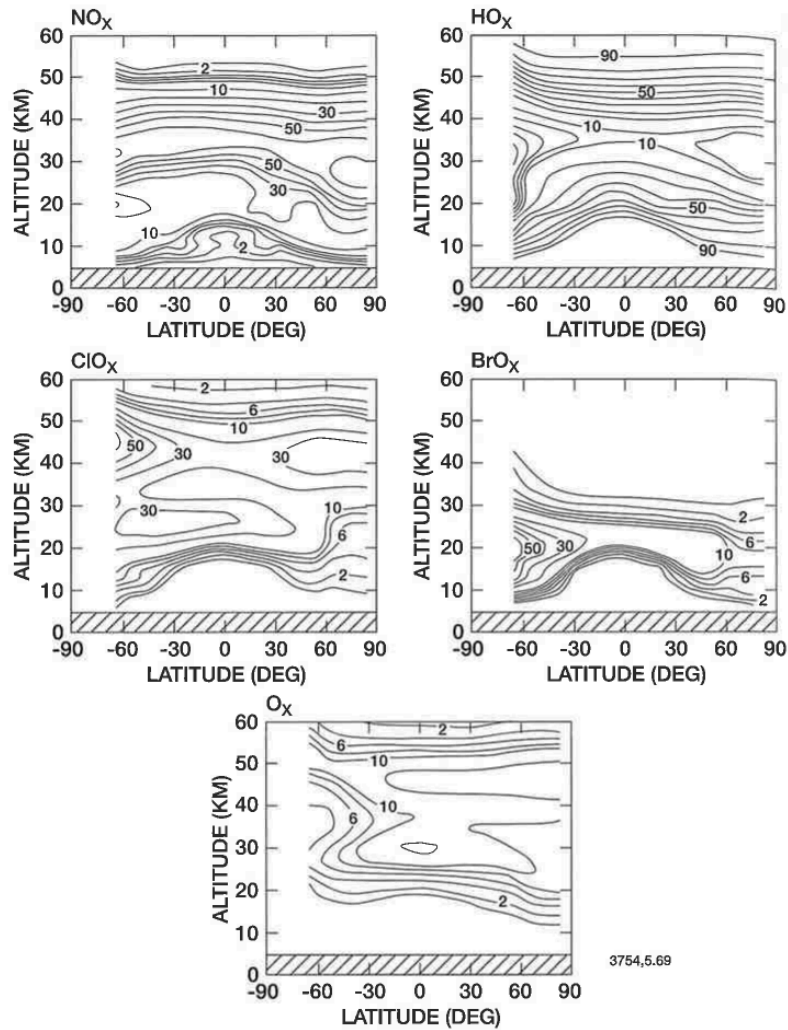
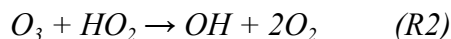
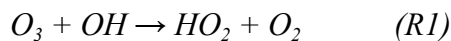


Figure 5.71. Relative contribution (percent) of various chemical families (NO_x , HO_x , ClO_x , BrO_x , and O_x) to the destruction of odd oxygen in the stratosphere in June (based on calculations by the NASA/Goddard 2-D model). Courtesy of C.H. Jackman, NASA.

We added more information in the manuscript, as follows:

“The spatial correlation with OH (Fig. 6) suggests that the ozone perturbation in the LMS is rather linked to lessened ozone destruction from the reaction $\text{O}_3 + \text{OH} \rightarrow \text{HO}_2 + \text{O}_2$, despite the higher springtime ozone abundance due to the Brewer–Dobson circulation (e.g. Cohen et al., 2018). At these altitudes, stratospheric ozone destruction is essentially caused by the HO_x (= $\text{OH} + \text{HO}_2$) catalytic cycle (with a contribution of $\sim 80\%$ in June, see Brasseur and Solomon, 2005), involving the reactions:



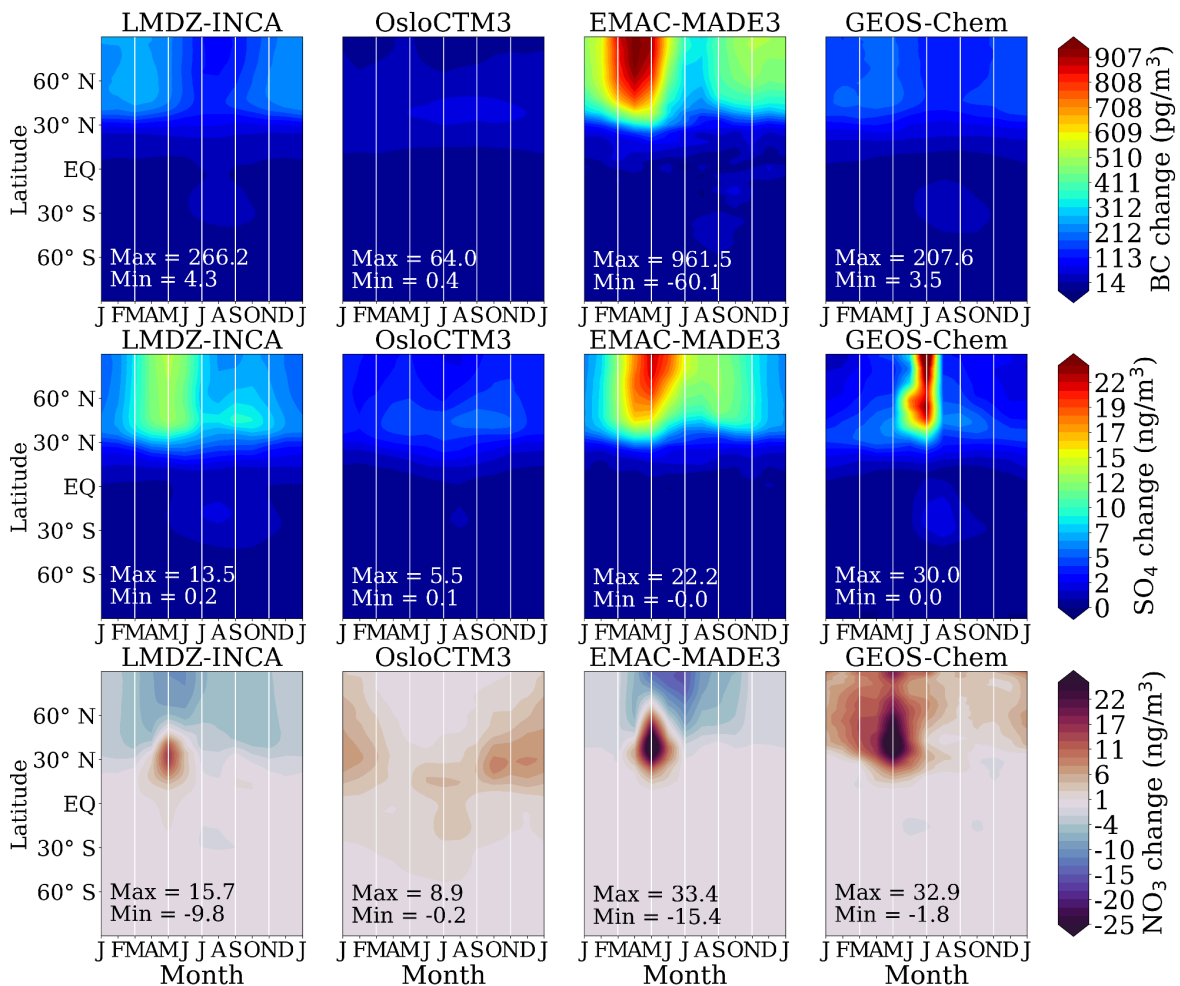
One can note that in the HO_x catalytic cycle, Reaction R2 is specific to the lowermost stratosphere, as $\text{O}(^1\text{D})$ is too rare to make the reaction $\text{O}(^1\text{D}) + \text{HO}_2 \rightarrow \text{OH} + \text{O}_2$ significant, contrary to the middle and upper stratosphere. Concerning the OH decrease in the LMS, we can relate it to the following reactions induced by the NO_x injection:





These four reactions explain that aircraft NO_x neutralizes the main ozone sink in this region. It is illustrated by enhanced NO_x levels extending into the polar LMS during the same season, and with HNO_3 increasing substantially (shown in Fig. S4 in Supplement). As the primary pollutants emitted mainly in the midlatitudes have their response extending into the LMS in winter as well (BC in Fig. 8, next section; SO_2 in Fig. S5, in Supplement), the wintertime confinement of the NO_x response in the midlatitudes cannot be due to transport only, which suggests a particularly short NO_x chemical lifetime compared to the poleward transport duration. During spring – summer however, part of HNO_3 is photolyzed back to NO_x , which can explain the northward extension of the NO_x response after the polar night.”

L.482: “The seasonality is the same as for ozone” – The seasonality here seems to be for sulfate aerosols. What about the seasonality for black carbon and nitrate aerosols? Are Jan and Apr-Jun as relevant as for ozone and sulfate aerosols?



The figure above shows Hovmöller diagrams for the three aerosol species, in the UTLS. The seasonality for SO_4 is different with GEOS-Chem, with a sharp but strong June-July maximum (especially July). Taking that into account, we had to slightly change our

description. For BC, the seasonality is very different than for ozone, as expected since BC is a primary species. But this kind of analysis is beyond the scope of the study, and we do not expect that the months we chose are relevant for every species. We clarified it. As to nitrate, the seasonality is commented on in the next paragraph. We readapted the paragraph as follows:

“Four models investigated the impact of aircraft emissions on aerosols. The LMDZ-INCA and OsloCTM3 models share the common protocol, GEOS-Chem follows a similar set-up (Quadros et al., 2025), while the EMAC-aer model is represented by the output from Righi et al. (2023) with a substantially different simulation set-up (see Section 2.2). The contribution of aviation to atmospheric aerosols is shown in Figs. 8–10. It is shown for the same months as for gaseous species in a purpose of simplicity, though it is not optimal for every species. It is worth reminding that the EMAC-aer model is more accurate (as it is equipped with a detailed two-moment aerosol microphysical scheme) than the other models (characterized by simpler aerosol representations). Also, the EMAC-aer model is not used with a QCTM mode, hence the existence of negative values in BC and SO₄ due to changes in dynamics and physical processes. For these two species, the models generally show a maximum in the UTLS as for NO_x in terms of zonal cross sections. OsloCTM3 has a much weaker response in the UTLS. For BC, all the models exhibit a local maximum at the mid-latitude surface, due to the take-off and landing phases (in Northeast America, Europe, and East Asia), and to subsidence. The LMDZ-INCA and GEOS-Chem models show similar responses, with a maximum in April. All the models except OsloCTM3 show a summertime minimum for BC. The sulfate perturbation reaches its maximum at high latitudes in May with LMDZ-INCA and EMAC-aer, and in July with GEOS-Chem (shown in Fig. S7 in Supplement). The SO₄ seasonality is similar with ozone, as photochemistry increases and promotes further the conversion of sulfur dioxide (SO₂) into SO₄, thus enhancing the formation of sulfate aerosol, as explained in Terrenoire et al. (2022) and Prashanth et al. (2022). For both BC and SO₄, the EMAC-aer model has a much stronger response, in the UTLS but also in the whole free troposphere. The maximum shifts from the lower free troposphere in winter up to the UTLS in summer. Most differences between EMAC-aer and the other models are consistent as EMAC-aer is the only model including the Aitken mode in the aerosol size distribution of aircraft emissions, and with an important proportion (91 % of emitted soot, and of primary sulfate particles, as supported by observations: Petzold et al., 1999; Mahnke et al., 2024), except the sulfate maximum in GEOS-Chem (July) that reaches higher levels than EMAC-aer.”

L.523: “fewer” – Do the authors mean “fewer than for the gas-phase”?

Yes, it is clarified now.

L.533: “surface concentrations” – Could the authors provide a figure/map in the supplement for the surface concentrations? How surface is defined in the models?

Accordingly with Referee 2's suggestion, these new figures have been added in Supplement (Figures S11–S13), and we filtered the “non-significant” gridcells for EMAC-aer, which removed some very high values mentioned in the preprint (for BC and NO₃). The corresponding paragraph has been modified as follows:

“It is worth mentioning the aviation impact on surface concentrations as represented by these models, shown in Figs. S11–S13. The surface is represented by the lowermost level (up to ~60 m, 70 m, 20 m, and 130 m for EMAC-aer, LMDZ-INCA, OsloCTM3, and GEOS-Chem respectively). As in Righi et al. (2023), we performed a t-test on EMAC-aer output in order to assess the significance of the perturbation against the interannual variability. We choose to show only the gridcells where the perturbation is characterized by a p-value lesser than 0.05, consistently with Righi et al. (2023). On average in the period of interest, the BC seasonal maxima at the surface are located at the eastern and western coasts of the US, in Europe, and, to a lesser extent, in East Asia (Fig. S11). These maxima are generally at ~1 ng m⁻³ except for EMAC-aer, which maxima reach ~ 4 ng m⁻³ in the eastern US, Europe and around the Mediterranean basin. The increase in the other two aerosol compounds is more significant. SO₄ perturbation takes place in western Europe, US, and the subsidence regions as North Africa–Middle East with a summertime average of ~35–45 ng m⁻³ (70–90 ng m⁻³ for EMAC-aer). For NO₃, the impact is generally stronger in winter in North America, western Europe, South Asia, and East Asia. The latter reaches wintertime NO₃ perturbations of 100–460 ng m⁻³. As for the UTLS, the differences between EMAC-aer and the other models are large, which suggests an important sensitivity of both climate and air quality impacts to the size of emitted aerosols (as discussed in Gettelman and Chen, 2013; Righi et al., 2013), and highlights the need for another model intercomparison with a more accurate aerosol parameterization in the model ensemble.”

L.639: “not modeled here” – Could the authors explain why they did not model “contrail cirrus formation”?

The inclusion of process-based contrail cirrus parametrisations in global climate models is a relatively new development, only available in a couple of models, in configurations specific to this purpose (Lee et al., 2021). The models and versions we use do not have the capability of quantifying the contrail cirrus effects. This has been updated in the text: *“Lastly, the other major component of the ERF from aviation emissions is contrail cirrus formation, which could not be simulated within the current capability of the models used in this study. Lee et al. (2021) estimated...”*

L.686-688: This is an important statement and seems to deserve being highlighted in the abstract.

We agree with Referee 2, and we added this statement to the abstract, as follows:

“Estimates of the aerosol direct ERF are systematically negative and range between -6.5 and -17.8 mW m⁻², compensating most of the net NO_x ERF though with noticeable intermodel differences arising from the diversity in aerosol parameterizations.”