Dear Editor, dear Reviewers,

We sincerely thank the reviewer for the careful evaluation of our manuscript, the encouraging assessment of its scientific quality, and the constructive comments and suggestions. We particularly appreciate the recommendations concerning the early emergence of ozone responses to SAI, the interpretation of the Northern Hemisphere ozone changes, and the definition of the integrated ozone column. We address these issues through additional analyses, clarification of the methodological scope, and revision of the relevant manuscript sections. The main changes introduced in the revised manuscript are summarized below:

- Clarification of the ozone column definition: It has been specified that the ozone column examined in this study denotes the partial ozone column integrated between 10 and 52 km, rather than the total ozone column. Terminology has been revised consistently across the manuscript, including figures, captions, and tables.

- Improved consistency of figure units: Panel (c) of Fig. 1 has been revised to display ozone retrieval uncertainties in molecular number density (molecules m^{-3}) instead of ppmv, thus ensuring consistency with the ozone concentration fields in Fig. 4.

- Concise reporting of results: The sections describing the results (Sects. 3.1 and 3.2) were made concise by removing repetitive numerical statements. A summary table (Table 1) has been created displaying some of the most informative global and seasonal PR-PO comparisons. This now complements the regional diagnostics (Table 2).

- Clarification of the study's scope and limitations: The discussion of the CAIRT-like observing concept, the choice of the hypothetical SAI scenario, and the distinction between reproducing the modelled long-term ozone response and detecting its initial emergence from natural variability have been strengthened. The discussion of complementary observations of other stratospheric species and of future applications to alternative multi-model SAI scenarios has also been expanded.

Please find more details and a point-by-point reply to the Reviewers' specific comments in the following (Reviewers' comments are in black and our replies in blue). We refer to specific text sections pointing at specific lines intervals (Lxxx-yyy, with xxx and yyy the starting and ending lines) of the revised manuscript.

Thank you very much,

Mona Kosary on behalf of all co-authors

Reviewer #1

This paper tests whether satellite mission concept CAIRT could actually detect the ozone changes due to stratospheric aerosol injection (SAI), if such a mission ever flew during a deployment. CAIRT is a mid-infrared limb-sounding mission concept that ESA considered for its Earth Explorer 11 program but ultimately (and sadly) didn't select. The authors use a set of synthetic "pseudo-observations" that include the retrieval noise, smoothing, and bias a real CAIRT-like instrument would introduce, including the extra effect of enhanced stratospheric aerosol loading on the retrievals themselves, and ask if those pseudo-observations preserve the model's simulated SAI-induced ozone signal, or if the instrument limitations would wash out the signal. Overall the paper's central result looks well-supported by the quantitative comparisons throughout, and the paper is perfect for this journal.

I have some mostly minor comments that need to be addressed before recommending publications.

Thank you for your kind words and your comments and suggestions for improvement. Please find in the following a point-by-point reply to your major and minor comments.

Major comments

1) I would frame the discussion a bit better in the broader context of literature around time of emergence of a climate signal from natural variability in the context of both climate change and ozone recovery. The authors do this in places already by discussing linear trends and calculating roughly how many years would be required to detect a signal, but they use a period in which the signal itself is already strong. Of much more interest could be to ask whether, if CAIRT was launched contextually with the SAI deployment, the mission would be able to detect the initial signal, and how long would that take.

We thank the Reviewer for this important and constructive comment. We agree that the early emergence of SAI-induced ozone changes from natural atmospheric variability is an important issue, particularly when considering an observing system operating from the beginning of a potential SAI deployment. The primary objective of our study is to assess the capability of CAIRT-like pseudo-observations to reproduce and quantitatively characterise the ozone response simulated by CESM2(WACCM6). We acknowledge that the approximate four-year averaging estimate presented in Sect. 3.3 is based on the established Antarctic ozone response over the full simulation period. Therefore, this estimate should not be interpreted as the time required to detect the initial ozone perturbation following the start of SO_2 injection.

We agree that identifying the initial SAI-induced ozone response may be particularly challenging within the nominal five-year lifetime of a single CAIRT-like mission, as the ozone perturbation may initially be small compared with natural variability. Our present analysis does not establish a specific time of emergence for this initial response. However, we would like to emphasize that detecting an SAI deployment does not necessarily require the immediate detection of its effects on ozone. A CAIRT-like observing system would also provide simultaneous measurements of other atmospheric constituents, including SO_2 and sulfate aerosols. As demonstrated in our previous study (Sellitto et al., 2026), CAIRT-like infrared limb observations could detect relatively small SO_2 injections and monitor the subsequent aerosol perturbations. These observations would provide valuable information on the occurrence and early evolution of SAI interventions, even when the associated ozone response remains difficult to distinguish from natural variability. We have clarified this distinction in the revised Introduction and Conclusions, highlighting the complementary roles of SO_2, sulfate aerosol, and ozone observations in monitoring potential SAI deployments. In addition, a complementary study entitled *“Monitoring the early multi-species stratospheric response to sulfur dioxide injection with CAIRT-like infrared limb sounding”* is currently being prepared for submission. This study investigates the capability of CAIRT-like observations to detect early changes in stratospheric composition during the first two years following SAI deployment, including the responses of ozone-related species such as NO_2 and N_2O_5. These species may provide complementary indicators of SAI-induced chemical perturbations before a clear ozone-column response can be distinguished from natural variability. We have now mentioned this ongoing work in the Introduction (132-134) and Conclusions (L547-549) of the revised manuscript.

Contextually to this comment, I want to note that while the paper is very strong and scientifically robust, it is a bit of a shame the authors selected to use a rather silly and unfeasible SAI scenario, especially when many multi-model results exist (i.e. both GeoMIP simulations and the new CCMI runs with SAI are out there, and have the same kind of ozone information this paper uses, see Jörimann et al. (2026) and Bednarz et al. (2026)), that have the added perks of having 1) more serious, believable scenarios compared to the Overshoot one, which would fit better with the “realistic SAI scenarios” the authors mention in line 159

and 2) multiple models which would enable the authors to better talk about detection/emergence within the context of different simulated ozone responses. Even if the authors still only looked at CESM2-WACCM6 here, they would have better enabled follow-up studies where someone else could have looked at this same question in other models within standardized scenarios.

Regarding the selection of the SAI scenario, we agree that the SSP5-3.4-OS overshoot experiment represents a specific hypothetical deployment pathway and should not be considered representative of all possible future SAI strategies – not one of the most feasibles either. We thank the Reviewer for suggesting the recent multi-model studies by Jörimann et al. (2026) and Bednarz et al. (2026). We agree that these studies provide an important framework for investigating the dependence of the simulated ozone response on the model and SAI strategy, and its detectability. We have therefore included these references in the revised Introduction and clarified our motivation for selecting the SSP5-3.4-OS experiment in Sect. 2.2, and mentioned that future work is needed to assess the detectability of ozone impacts of SAI in other and more feasible scenarios. In particular, we have expanded the Conclusions to discuss future applications of the present CAIRT-like pseudo-observation framework to alternative, standardized multi-model SAI experiments. Such studies would allow the observational capabilities investigated here to be assessed across different modelled ozone responses and more moderate deployment scenarios.

These modifications are included in the revised introduction (L105-110), Sect. 2.2 (L213-217), and Conclusions (L550-552).

2) One potential discrepancy: the abstract's point (c) claims "a moderate ozone increase in winter and spring at northern hemispheric mid and high latitudes." But Table 1 shows (0–60°N) increasing while (60–82°N) actually decreases slightly in both the annual mean (−0.74 DU) and in March specifically (−0.73 DU). There's no NH high-latitude winter analysis shown anywhere I could find to support the "winter" part of the claim. As written, the abstract overstates/misattributes the high-latitude finding, and it should probably be split into "increase at NH mid-latitudes" versus "weak, non-robust decrease at NH high latitudes," or supported with an added winter-specific analysis. Related to this, the abstract's causal mechanism — "associated with altered transport and weakened subtropical jets" should be better discussed and referenced (there are plenty of works out there discussing this) in the paper.

We thank the Reviewer for pointing out this inconsistency. We agree that the statement in the Abstract was too general and did not accurately reflect the different ozone responses found across Northern Hemisphere latitude bands.

As reported in Table 2 in revised version of manuscript, the annual-mean geoengineering–baseline difference in the PR is positive over 0–60°N (+1.67 DU), whereas it is slightly negative over 60–82°N (−0.74 DU). The corresponding March anomaly at northern high latitudes is also slightly negative (−0.73 DU). However, these high-latitude differences are small compared with the interannual variability (6.54 DU for the annual analysis and approximately 10.51 DU for March), indicating that the modelled Arctic response is weak and not robustly distinguishable from variability. We have therefore corrected the Abstract to distinguish between the modest annual-mean ozone increase at northern low and middle latitudes and the weak, slightly negative response at high northern latitudes. We have also removed the unsupported reference to a winter and spring ozone increase at northern high latitudes.

Regarding the proposed dynamical explanation, we agree that the statement about altered transport and weakened subtropical jets was not sufficiently supported by the analyses presented in our study. While the influence of SAI on stratospheric chemistry and circulation is discussed in the Introduction, with references to previous modelling studies, our paper does not really focus on the attribution of the simulated ozone response to specific dynamical mechanisms (rather, it is about its detectability). For this reason, we think

safer to remove the specific attribution to weakened subtropical jets from the Abstract. The Conclusions now state explicitly that process attribution is outside the scope of this observing-system assessment (L517–518). We have also added a clarification in Sect. 3.2, highlighting the spatially heterogeneous Northern Hemisphere response and the limited robustness of the Arctic anomalies.

The corresponding modifications are included in the revised Abstract (L30-32), Sect. 3.2 (L411-416 and L424-432), and Conclusions (L515-518).

3) The authors define the TOC in line 320-321 as the integral from 10 to 52 km – I found this a bit weird, as that is a partial, stratosphere-only column, not the conventional total-atmosphere column. But the term "Total Ozone Column" is already used in Figure 2 and its caption (p. 9), well before this definition appears, without stating the integration bounds there. Since typical satellite/ground-based "total column ozone" (for instance, from OMI) products integrate the full atmosphere including the troposphere, this could confuse readers comparing the paper's DU values to standard products. I suggest moving the definition earlier, and justifying it a bit better in the context of other products.

We agree that the terminology "total ozone column" was not appropriate for our analysis, as the integration is restricted to the altitude range between 10 and 52 km and therefore does not represent the conventional total atmospheric ozone column. The selected integration range corresponds to the vertical domain considered in our CAIRT-like ozone analysis. By applying the same integration limits to both PR and PO fields, we ensure a consistent comparison of the modelled SAI-induced ozone response and its representation in the pseudo-observations. To clarify this point, we have moved the definition of the vertically integrated ozone quantity to Sect. 2.4, before its first appearance in Fig. 2, and explicitly distinguished it from conventional total atmospheric ozone column products, which include the contribution from altitudes below 10 km. We have also replaced the terminology "total ozone column (TOC)" with " partial ozone column (OC)" throughout the manuscript, including the relevant figures, captions, and tables, and explicitly stated the 10–52 km integration limits.

The definition added in Sect. 2.4 (L264-270) reads:

"Throughout this study, the vertically integrated ozone quantity is defined as the partial ozone column (here defined as “partial OC”), between 10 and 52 km altitude, calculated by integrating the ozone number density profiles and expressed in Dobson units (DU). This altitude range corresponds to the vertical domain considered in the present CAIRT-like analysis. The same integration limits are applied to both pseudo-reality (PR) and pseudo-observation (PO) fields, ensuring consistent comparisons of the modelled and pseudo-observed ozone responses to SAI. Unlike conventional total ozone column products, e.g. from nadir satellite or ground-based observations, which account for the entire atmospheric column, our calculation excludes ozone below 10 km."

Minor comments

1) **Abstract, line 20-21**: "to assess how a future mid-infrared limb-emission sounding observing system would characterise its modelled ozone response": "its" is ambiguous (does it relate to the system's response? the SAI scenario's response?). Maybe "...would characterise the modelled ozone response."

Agreed. We removed “its” and changed to “would characterise modelled ozone response to SAI” to clarify that the reference is to the modelled ozone response to SAI.

2) **Line 122-124**: "which is investigated in details in this study" should be "in detail"

Thank you, corrected to “in detail”.

3) **Line 155**: "in the hypothesis of a series of such instruments proving consistent observations", "proving" should almost certainly be "providing." "In the hypothesis of" also reads weird; "under the assumption that a series of such instruments provides..." would be clearer.

We agree and have changed the sentence for clarity, including an explicit reference to the nominal single-mission lifetime. We changed the revised manuscript into “The present work extends the ozone analysis beyond the nominal lifetime of a single CAIRT mission, under the assumption that a series of CAIRT-like instruments could provide consistent observations suitable for decadal studies.”.

4) **Figure 2 caption**: "Global maps of the total ozone column averaged derived from model over the period 2033-2062" is a bit garbled. Suggest: "Global maps of the model-derived total ozone column, averaged over 2033–2062, for..."

Thank you for this correction. We have changed in the revised manuscript.

5) **Line 552-553** (Conclusions): "the recent selection of NASA's STRIVE mission demonstrate that..." should be "demonstrates."

We corrected to “demonstrates”.

The reference list includes at least one entry (Budyko, 1977) that I could not find cited anywhere in the text.

We checked the manuscript and removed the unused Budyko (1977) reference from the bibliography.

Figure 4 uses molecules/m³ (nO_3, $\times 10^{18}$ molec/m³), but the text discussing the same quantity using mol m^{-3}. Clearly the authors use mol as a shorthand for molecules, not moles, but just be clear about it.

We have corrected the units in the relevant Results paragraph and Figure 4 to “molec m^{-3}”.

Review2

General Comments

1) This goal of this paper is to demonstrate that a future mid-infrared limb-emission sounding instrument, like CAIRT, has the precision/resolution to determine differences in stratospheric ozone which may occur due to a future geoengineering scenario between 2035 and 2060. The “data” used for this comparison comes from CESM2/WACCAM model projections under scenarios of business as usual and a geoengineering injection rate of 10 Tg SO2/yr. The differences in ozone between these runs is on the order of plus/minus a few percent to over 10% depending on altitude, latitude, and hemisphere, with generally ozone loss in the southern hemisphere. In addition to the direct model output of ozone, pseudo reality (PR) in both scenarios, the authors also calculate pseudo observations (PO), i.e. the observations a mid-infrared sounding instrument would make, were it available through the two scenario periods. If the differences, geoengineering to natural (del PR), are greater than del PR/PO, the differences between PR and PO, then

the limb sounding instrument would be capable of detecting del(PR) proving the authors thesis that a CAIRT like instrument is needed to assess future possible del(PRs).

We thank the Reviewer for summarizing the objective and methodology of our study. We would like to clarify that the simulations analysed here cover the period 2033–2062, with SO_2 injections beginning in January 2034. The injection rate is not constant at 10 Tg SO_2 yr^{-1}, but varies over time, reaching approximately 10 Tg SO_2 yr^{-1} around 2050, as described in Sect. 2.2. We agree that comparing the modelled geoengineering–baseline ozone response with the differences introduced by the pseudo-observation processing is a central element of our assessment of the capability of such instrument to monitor the impacts of this selected SAI scenario on the stratospheric ozone layer.

However, our objective extends beyond a direct comparison of these magnitudes. In particular, we investigate whether CAIRT-like pseudo-observations preserve the spatial distribution, regional variability, seasonal characteristics, and long-term evolution of the modelled ozone response, including changes in the timing of Antarctic ozone recovery. We emphasize that the study assesses the ability of CAIRT-like observations to reproduce and characterise these complex patterns of the ozone perturbations, in the selected pseudo-reality scenario. This is fundamentally distinct from establishing the statistical detectability or attribution of the initial SAI-induced ozone response against natural atmospheric variability and, in our opinion, justifies the different analyses presented here.

2) Fig. 1, the key figure in this paper, shows that the 1 sigma uncertainty, in the stratosphere, ranges from 0.05-0.15 ppmv for a limb sounding instrument. For a typical stratospheric ozone profile this corresponds to uncertainties of 0.5 – 3%. While the upper end of this difference approaches del(PR), the following figures in the paper demonstrate, over and over again, that del PR/PO is less than 1% in both scenarios, proving the authors point from the beginning. Never the less, the authors proceed to describe the differences globally, regionally, and seasonally. Every section proceeds the same way showing that del PR/PO is much less than del PR, and then list the differences in the text of the paper. This makes for very dry reading. If the value of these differences are important, in all the different data slices, they should be put into a table where they could be seen at a glance. Meanwhile the interesting fluctuations and differences in the two PR scenarios are not mentioned at all.

First of all, we accept the criticism of the Reviewer about the fact that the repeated presentation of numerical PR–PO comparisons might reduce the readability of the Results section. We have therefore shortened the discussion in Sect. 3.1(L324-332 and L344-352), retaining the main physical characteristics of the simulated ozone response and the most informative observing-system performance diagnostics, thus suppressing most of the repeated text. The global mean scenario differences and corresponding double difference are now summarized in Table 1 in revised version of manuscript, allowing a direct quantitative comparison without repeating the individual values throughout the text. We have also placed greater emphasis on the vertical, latitudinal, and temporal structure of the ozone response in the PR fields, while retaining the assessment of how accurately these features are reproduced by CAIRT-like PO.

In Sect. 3.2 (L376-404 and L411-432), we have also reorganized the regional discussion to distinguish more clearly between the physical characteristics of the modelled ozone response and the additional deviations introduced by the pseudo-observation processing. Annual regional statistics are summarized in Table 2 in revised version of manuscript, while selected global and seasonal PR–PO comparisons are consolidated in the new table. This allows us to reduce repeated numerical descriptions while retaining the key results concerning Antarctic ozone depletion, changes in the timing of ozone recovery, and the substantial interannual variability of the Arctic response. In particular, we have clarified that the small long-term mean Arctic ozone anomaly does not imply the absence of appreciable differences between the scenarios in individual years.

3) It is not clear why this paper should be published as the point is obvious from the beginning and CAIRT was not selected. Should the editor choose to proceed anyway, then the paper should be significantly shortened and homogenized.

We thank the Reviewer for raising this point. We respectfully disagree that the scientific relevance of the study depends on whether CAIRT was selected as the ESA Earth Explorer 11 mission.

The objective of this work is to assess the capabilities of a class of observing systems based on mid-infrared limb-imaging Fourier-transform spectroscopy and tomographic retrieval, rather than to evaluate a specific funded satellite mission. Although imaging Fourier-transform spectrometers have already been used in space in nadir-viewing geometry (e.g. GIIRS on FY-4 and MTG-IRS), their application to limb imaging with tomographic retrieval represents a novel measurement concept, building on the heritage of the airborne GLORIA instrument. Therefore, despite CAIRT not being selected within the EE11 programme, the proposed observing concept remains scientifically relevant for monitoring stratospheric composition, including under hypothetical SAI deployment scenarios. The recent selection of NASA's STRIVE mission further highlights the continuing relevance of limb-infrared observations.

We also acknowledge that the retrieval uncertainties presented in Fig. 1 already indicate favourable measurement sensitivity to ozone. However, these diagnostics alone do not establish how the full pseudo-observation processing, e.g. including retrieval-related uncertainties, spatial smoothing, and the impact of enhanced aerosol conditions on ozone retrievals, affects the representation of the modelled SAI-induced ozone response.

Our study therefore extends this initial point performance characterization by quantitatively evaluating the preservation of the spatial distribution, seasonal variability, and long-term evolution of the ozone response, particularly the delayed Antarctic ozone recovery simulated under the investigated SAI scenario. We think that this assessment requires the analyses presented in the figures.

In any case, we agree with the Reviewer that the manuscript might benefit from a more concise presentation. As discussed in our response to the previous comment, we have reduced repeated numerical descriptions in the Results and the inherent text, and consolidated selected diagnostics into summary tables. We have also clarified the scope of the study as an assessment of a CAIRT-like observing concept, without implying that the investigation represents a confirmed future mission or establishes the statistical detectability of the initial SAI-induced ozone response. We believe that these revisions clarify the scientific contribution of the work and its relevance to future mid-infrared limb-emission observing systems.

4) All the endless text quoting numbers should be reduced to tables. All the figures should be homogenized as much as possible. Don't show uncertainty in mixing ratio and then the latitude altitude ozone field in molecular concentration. Use one or the other units for both. Better would be to show uncertainty in percent, and the same for del(PR) throughout. Notice how little difference there is in del(PR/PO) throughout for both scenarios, see Figs. 4, 5, 7-10. At least in Figs. 7-10 the authors add percent difference. In Fig. 4 del(PR/PO) is so small the difference scale has to be amplified.

We thank the Reviewer for the suggestion to improve the consistency of units and graphical presentation across the figures. We agree that presenting ozone retrieval uncertainties in mixing ratio (ppmv) in Fig. 1, while showing ozone number density (molecules m^{-3}) in Fig. 4, may complicate their direct comparison. To address this concern, we have revised panel (c) of Fig. 1 to express the ozone retrieval uncertainties in molecular number density (10^{18} molecules m^{-3}), consistently with Fig. 4. This modification facilitates comparison between the expected retrieval uncertainties and the magnitude of the modelled ozone perturbations. Figures 5–10 present the vertically integrated ozone columns in Dobson Units (DU), which

are appropriate for evaluating changes in the ozone column (i.e., not at a specific altitude) and their temporal evolution. Furthermore, Figs. 7–10 already include relative geoengineering–baseline differences in percent, allowing the magnitude of the modelled ozone response to be evaluated relative to the corresponding baseline. Although expressing all differences in percent could facilitate some comparisons, we believe that retaining absolute quantities is also important for the readers, particularly for interpreting the magnitude and vertical distribution of ozone perturbations. Relative differences can become disproportionately large in regions where background ozone concentrations are relatively low, even when the corresponding absolute changes are small. Also, we don't think it would be particularly useful, in this context, to express the point retrieval uncertainties of Fig. 1c in percent (we think it makes much more sense to get information on the absolute uncertainty of the parameter that is actually measured), and then a ready comparison of the differences of the following figures calls for retaining the absolute differences. Regarding Fig. 4, we acknowledge that the PR–PO residuals are considerably smaller than the geoengineering–baseline differences. For this reason, different color scales are used to make the spatial structure of the residuals visible. These scales are explicitly indicated in the figure and the caption, and we have clarified this choice in the revised manuscript.

Specific Comments

Here are a few additional comments/questions.

1) 321 Isn't it Fig.5 showing the latitude-time evolution?

We corrected to "Figure 5".

2) 246 isn't 24 DU closer to 10% than 5%?

We thank the Reviewer for pointing out this inconsistency. After checking the data used to generate Fig. 2, we confirmed that the maximum absolute difference between the two scenarios is approximately 20 DU, rather than 24 DU as originally stated. The maximum baseline partial ozone column is approximately 356 DU, giving a ratio of about 5.6%, consistent with our initial approximate indication of 5%. However, since these maxima do not necessarily occur at the same location, this ratio should not be interpreted as the maximum local relative difference. We have therefore corrected the absolute difference and clarified the wording in the revised manuscript.

3) Fig. 8 What is causing the rather large annual to bi-annual oscillations in both the baseline and geoengineering cases? Why aren't these oscillations correlated more closely between baseline and geoengineering? Why are they completely absent in the October means, particularly after 2050? Similar oscillations also occur in Fig. 9 where the anti-correlations are even more pronounced.

We thank the Reviewer for highlighting the short-term fluctuations in the Antarctic and Arctic ozone time series and the differences between the baseline and geoengineering scenarios. We clarify that the different appearance of the time series is partly explained by the temporal smoothing applied for visualization. Figures 7 and 9 present monthly regional-mean ozone columns smoothed with a 12-month running mean, whereas Figs. 8 and 10 show October and March values, respectively, smoothed with a five-year running mean, as indicated in the caption. The stronger smoothing substantially attenuates annual-to-biennial fluctuations. Therefore, their reduced visibility in Fig. 8, particularly after 2050, should not be interpreted as evidence that interannual variability is absent from the underlying October data.

4) 424 … hardly discernable…? That is not the case. There are easily differences of 10 DU or more discernable. It is just that with 12 month averaging, and unexplained differences in the short term variations

between the baseline and geoengineering, that the average differences aren't as large as in the Antarctic. Still the positive differences are 50% and the negative differences 25-50% of the Antarctic differences. These differences are borne out by Fig. 10 where the influence of the annual averaging is diminished.

We thank the Reviewer for this clarification and agree that the expression "hardly discernible" was misleading. Although the long-term mean Arctic ozone response is considerably smaller than that observed over Antarctica, individual years exhibit substantial positive and negative geoengineering–baseline differences. In the PR simulations, the Arctic annual-mean anomalies range from −10.43 to +13.13 DU, while the long-term mean difference is only −0.74 DU. The March analysis shows even larger interannual variations, ranging from −24.19 to +28.35 DU. We therefore agree that the Arctic differences are clearly discernible in individual years. Our original intention was to emphasize the relatively small and less persistent long-term mean response rather than the absence of short-term differences. We also clarify that the 12-month running mean applied in Fig. 9 attenuates short-term variations for visualization, but does not explain the small multi-decadal mean difference. The latter reflects the compensation between positive and negative anomalies over the analysed period. We have revised Sect. 3.2 (L411-416 and L424-432) accordingly, replacing the expression "hardly discernible" and emphasizing the distinction between substantial interannual variability and the small long-term mean Arctic ozone response.