Replies to Referee 1:

1. Page 4, Figure 1: The instrument datasets listed for creating GOZCARDS are SAGE-I, SAGE-II, HALOE, and Aura MLS. However, UARS MLS and ACE-FTS were also included in the GOZCARDS ozone database according to Lucien et al. (2015) (reference listed in Table 3). Please verify whether these two instruments were unintentionally omitted from Figure 1 or whether they are no longer included in the latest version of the dataset.

   We verified that the current GOZCARDS version uses SAGE-I, SAGE-II, HALOE and Aura MLS and not the UARS MLS or ACE-FTS data. This was confirmed directly by Lucien Froidevaux.

2. Page 5, Figure 2: OMPS-NP data have been used in several merged datasets, such as SBUV/NASA (MOD), SBUV/NOAA (COH), and MSR-2. For clarification, more detailed information should be provided to distinguish the different OMPS-NP satellites. For example, it would be better to replace "OMPS-NP" with "OMPS-NP (Suomi-NPP)" for the SBUV/NASA (MOD) dataset. More detailed information should also be provided for the "OMPS" instrument used in MSR-2.

   Thank you for pointing this out. The specific OMPS satellite information was added to Fig. 2.

3. Page 6, Table 1: Similar to comment #2, more detailed information is needed for "OMPS" in the instrument column for MSR-2. In addition, for NOAA/SBUV (COH), the OMPS-NP instrument from NOAA-20 is missing in Table 1, although it appears in Figure 2.

   We added the missing NOAA-20 instrument to Table 1.

4. Page 8, Figure 8: It is not clear from either the figure caption or the text in Section 3.1 what "reference" value was used to define the differences before calculating the trends in those differences. In other words, was the reference based on MLS Level-3 or Level-2 data (e.g., by first converting individual profiles)? This information would help readers better interpret the sign of the trends in the differences and the impact of coordinate conversion on the derived ozone trends.

   Page 8, Figure 3 shows the trend of the differences (post-gridded - pre-gridded) between MLS data converted at L3 (conversion after gridding) and L2 (conversion before gridding). We clarified this in the revised version.
   "Zonal mean decadal trends of differences between MLS L3 data converted after and before gridding (post-gridded - pre-gridded) in % per decade from 2005 to 2022 at altitudes of 25, 30, 35 and 40~km (panels a to d)."

5. Lines 153-154, "It ranges from 0.5 to 0.75%/decade between 60^0 S and 60^0 N and above 20km". This statement is applicable for 2000 to 2024 period, but not for 1985 to 1995 period (Figure 4a). The sentence should be revised for clarity.

   We revised the sentence for clarity and added additional sentences to explain the

differences between both periods.
"In the period between 1985 through 1995, the uncertainties show higher values from close to the tropopause up to 25 km for latitudes above 50° N/50° S. Above this altitude and in the zonal band between these latitudes, the uncertainties are comparable in magnitude with those from the later period"

6. Lines 156-157, "and the results are shown for the period after 1996 (ODS peak) and 2000 in Figure 5." This sentence implies that the total ozone data periods shown in Figure 5 are 1996–2024 and 2000–2024 (red and blue lines). However, the Figure 5 caption indicates "until 1995 and since 2000" (i.e., 1985–1995 and 2000–2024). Please clarify the wording for consistency.

   We fixed the figure caption to correctly say "since 1996 and 2000", to mirror the text, stating now "results are shown for the period since 1996 (ODS peak) and 2000 in Fig. 5.".

7. Page 10, Equation (2), there are typos in the "inflection time" corresponding to the linear trend coefficients, β13(z,t)(t-t1), and β16(z,t)(t-t2). The terms t1, and t2 should be corrected to t0, and t1, respectively.

   We fixed the t1 and t2 terms, in order to agree with Eqs. (3), (4) and (5).

8. Lines 174-175, A special term, $L_{Gap}(t)$, is used for the 1996-1999 period due to "the sparse sampling of the merged ozone profile datasets". During this period, the main contributing satellite measurements came from SAGE II and HALOE (see Figure 1). Since there were no major volcanic eruptions (Figure A1) or extended periods of instrument problems during that time, could the authors elaborate further on the cause of sparse sampling, and did the sparse measurements mainly occur in specific latitude or altitude regions?

   We changed the text to better coincide with the SPARC LOTUS Report. The gap period was introduced because the trend inflection point varies in dependence on the altitude and latitude (Laine et al., 2014) making a selection of a single inflection point sub-optimal.
   "Note that the term with $L_{Gap}(t)$ does not contain a linear trend coefficient during the gap period (1996–1999). This is needed to reduce the influence of the selection of a turnaround point between ozone decline and recovery because the timing of the trend reversal varies with latitude and altitude (Laine et al. 2014). By introducing a transition period rather than a single inflection year, the regression model allows for a more gradual change between the two trend segments."

9. Page 14, Figure 6. The regression coefficients of the proxy terms are shown as functions of latitude and altitude, but there is no information regarding the statistical significance of those coefficients. It would be helpful to use hatching to indicate insignificant coefficients (similar to Figure 3). This would provide additional insight into the remarkably different impacts of the three volcanic aerosol proxies (Figure 6i, j, k). A related question concerns Figure 6k, the El Chichon aerosol proxy appears to result in ozone increases in the mid-

latitude lower stratosphere (below 20km). Is this feature an artifact of MLR fit or has it been reported in previous studies?

The significance of the regression coefficients was added to Fig. 6.
There are studies that show that El Chichon might contribute to an ozone increase (Chehade et al. 2014). We included the following text in the section about the additional proxy data and a note that the El Chichon regression coefficients are mostly not significant statistically.
" Volcanic eruptions have different impacts on stratospheric circulation, chemistry, and ozone distribution, depending on magnitude, season, and latitude of the event (Stolarski et al., 2006; Chehade et al., 2014; Weber et al., 2018, 2022). This variable response is illustrated in panels i-k of Fig. 6. The Australian wildfire (end of 2019) and HTHH had a stronger initial impact in the SH than in the NH because the volcanic aerosols (and stratospheric water vapour) were transported mostly in the SH resulting in an ozone loss (band of negative values in panel i). The Pinatubo eruption had a stronger impact in the NH (boreal winter) as illustrated by the band of negative coefficients from the tropics to the polar region (panel j, Aquila et al. (2014)). The dynamic response to the El Chichón eruption showed the opposite effect. Due to changes in the SH circulation the dynamical ozone increase exceeded the chemical ozone loss caused by the eruption (see panel k, Zerefos et al. (1994); Angell (1997)). However, nearly all El Chichon coefficients are not statistically significant at 95% confidence level."

10. Line 285, "SH trends are lower and remain at about +0.5DU/decade". The trend value appears to be a typo. It should be approximately 1 to 1.5 DU/decade (or ~0.5 %/decade).

    We fixed this mix up between DU/decade and %/decade. This was indeed 1.5 DU/decade and 0.5%/decade.

11. Line 354, "in % per decade Fig. A7) can can be compared". Please correct the typos, for example, "in % per decade (Fig. A7) can be compared".

    We corrected the typos as suggested.

12. Lines 359-360, "The larger variations in SH trend profiles, ….. is due to the smaller positive trends seen in GOZCADRS compared to all other merged datasets." The smaller positive GOZCARDS trends mainly occur above 25 km. Below 20 km the larger variations in trend profiles are mainly associated with SAGE-SCIA-OMPS and SAGE-OSIRIS-OMPS. For clarification, I recommend revising the sentence to specify the altitude range, for example, "The larger variations in SH trend profiles for altitudes **above 25 km**, …, is due to the smaller positive trends seen in GOZCARDS….".

    We added the suggested altitude range to the text.
    "The larger variations in SH trend profiles for altitudes above 25 km, seen in panel a) of Fig. 10 (see also Figs. A7-A9) is due to the smaller positive trends seen in GOZCARDS compared to all other merged datasets."

13. Lines 417-418, “For the stratospheric column calculations, data were screened out if more than three of the five layers above the tropopause had missed data”. Conventionally there should be no missing data above the tropopause when calculating stratospheric column ozone. Which five layers were used for screening criteria, and how were the missing data treated (e.g. interpolation) before calculating the stratospheric column ozone?

This was indeed confusing and we changed the wording and sentences in the manuscript as follows:
“The ozone profile datasets generally start at an altitude of 10 km (Table 1) which can be an issue when the tropopause is lower than 10 km, particularly at higher latitudes. For the stratospheric column calculations, data were screened out if the three lowest 1 km layers above the tropopause had missing data. In combination with the larger uncertainties of limb measurements below 20 km, this selection criterion causes larger spreads in the profile and stratospheric column trends. “
At higher latitudes, some ozone profile measurements contain missing values above the tropopause. To account for this issue, we applied a screening criterion based on the missing 1 km layers above the tropopause. Stratospheric columns were only calculated when at least 3 km above the tropopause was valid ozone data. Missing data inbetween two valid data points (if any at all) was handled by the merged datasets themselves. No interpolation was done or needed in this analysis.
Because missing values in these high-ozone layers could potentially affect the resulting column values and the derived ozone trends, we additionally compared our approach with the method described by Sofieva et al. (2026). Here we first calculate the ozone trends of the profile ozone datasets and then integrate the columns. The comparison of both layer and column ozone trends is presented in Figure A6. The results show that the two methods produce very similar trends, particularly for the Full Regression analysis, indicating that the treatment of missing values has only a minor impact on the overall trend estimates. Therefore, results obtained with both methods are provided.

Replies to Referee 2:

14. perhaps add in the title that trends are (predominantly) calculated from satellite ozone retrievals

We thought about different titles, including adding “satellite data”, but came to the conclusion, that it would make the title very long without conveying more information. We prefer to not change the title.

15. line 60 ends with two full stops.

We removed the second full stop.

16. lines 136-138: “At 25 km, a negative trend of around 1 % per decade between April and August in the northern hemisphere (NH) and between January and July in the southern hemisphere (SH) is noted. The significance of these trends gets smaller with altitude, showing the most significant trends at 25 km.” Give a possible reason/explanation for this finding.

Figure 3 shows the estimated trend uncertainty associated with the unit conversion. The enhanced significance around 25 km may be related to stronger vertical ozone gradients, which can amplify uncertainties arising from the conversion and gridding procedures. We added the following text:
“This is possibly because of stronger vertical ozone or temperature gradients near this altitude, which increase the uncertainties from the non-linear ozone unit conversion and vertical axis transformation.”

17. Section 3.2: can you make some (general) conclusions on the relative drifts of the individual merged datasets w.r.t. the MIM? This might be interesting to link to some “deviating” trends that were found for some of the datasets (e.g. the ones containing OSIRIS) in Fig. 13.

The datasets containing OSIRIS do have a stronger positive drift compared to the MIM, while the other four datasets have a slight negative drift. This is true for nearly all altitudes and latitudes, while Fig. 13 shows these differences mostly in the northern hemisphere. The drifts between datasets are identified to be the largest contributor to the spread in the derived trends from MLR.

18. In Figs. 4 and 5, you show the drift uncertainties for two different time periods, but you don’t mention why these different time periods are considered and if the drift uncertainties are different between those two time periods. If you include it, describe it!

We agree with the reviewer and added a short explanation why show two different drift uncertainties in Fig. 5 and how these drift uncertainties differ. The text was changed accordingly at line 156:
“The same procedure was applied to the total ozone datasets, and results are shown for

the period since 1996 (ODS peak) and 2000 in Fig. 5. We did not include the period before the ODS peak in this figure, because two of the total ozone datasets start around 1995 (GTO-ECV and GSG), leaving us with a limited selection of merged datasets and instruments before that year. Instead we show the relative drift uncertainty for two different start years (1996 and 2000), which were used in previous studies (Weber et al., 2022; Godin-Beekmann et al., 2022; Sofieva et al., 2026). Both relative drift uncertainties look very similar and show no significant differences. The drift uncertainties in the period 2000-2024 are slightly larger than in the period starting in 1996. When comparing the drift uncertainty for the total ozone datasets to the profile ozone datasets (Fig. 4), we see that the total ozone dataset drifts are overall lower and range between 0.1 and 0.5 % per decade..”

19. line 158: the highest total ozone drift uncertainties can be seen at high latitudes and in the tropics. Why is that? Again due to the higher uncertainties in the total ozone measurements there? Explain!

We added the following text after the line 158:
“**Higher drift uncertainties at higher latitudes are most probably due to the combination of higher uncertainties in ozone itself (higher solar zenith angles) and a larger natural variability in ozone. It is not clear why the drift uncertainties are enhanced in the tropics. Probably the low natural ozone variability masks the drifts to a lesser degree.**”

20. lines 238-245: I’m not very familiar with the concept of accumulating proxies over their dynamically active season and could not find how it is done in earlier studies (Weber et al., 2018, 2022). How is it done in practice? For example, a winter accumulated proxy is obtained by using the original Dec value as the Dec MM, the original Dec value + the original Jan value as the Jan MM, etc?

Yes, accumulated proxies are constructed as cumulative sums over their dynamically active season. The proxy value for a given month represents the sum of the monthly proxy values from the beginning of the season up to and including that month. This reflects the cumulative influence of dynamical transport on stratospheric ozone variability (Weber et al., 2011).

21. line 266: add a right bracket after Fig. A1

Done

22. lines 270-272: it’s a good idea to run the MLR model over the entire time period available, also when only calculation trends from 2000-2024. But I was wondering if the trends would differ significantly if only the 2000-2024 period would be regressed? Have you tried this sensitivity experiment?

In order to be able to compare results from our study to Sofieva et al. (2026) and for consistency with earlier studies e.g. Godin-Beekmann et al. (2022), we do the regression

over the full period, but showing only the results after 2000. The regression using the longer period is helpful for better constraining the proxy terms in the MLR. In general, the trends evaluated with the MLR using 2000-2024 are similar to post-2000 trends using the whole period, but the trend uncertainties are larger when a shorter time period is used.

23. Fig. 7: can you make the so-called thick lines thicker? They are hard to distinguish from the thin lines on my printout.

We increased the thickness of the lines a bit. From the later comments, it seems that the PDF reader might have an issue with some of our plots here, because we do use dashed and dotted lines in the different figures. We will use a different format to save the files. hopefully this will fix this and the other issues as well.

24. Fig. 8: The dataset lines are very hard to see.

Increased the thickness and alpha of the dataset lines.

25. Fig. 9, caption: the climatological thermal tropopauses are marked by full lines, not dashed lines. You might make those full lines thicker as well.

See our reply to comment 10.

26. Lines 360-361: any explanation for the smaller SH positive trends seen in GOZCARDS, which seem to get closer to the trends of the other data when the MLR is applied on deseasonalized ozone data?

We have not examined the smaller SH positive trends of GOZCARDS in detail, but we did reproduce the results of Sofieva et al. 2026, using deseasonalized ozone data. We suspect this difference comes from the seasonality that was either removed by the deseasonalization or not caught by the MLR.

27. Line 362 and caption Fig. 10: the joint-distribution uncertainties are marked by full lines (not dashed lines)

See reply to comment 10.

28. Fig. 12: can the thick lines be made thicker?

See reply to comment 10.

29. Lines 393-396 and 405-406: why are the trends from the datasets containing OSIRIS stronger than trends from other merged datasets? What’s wrong with OSIRIS? Are these datasets also drifting from the MIM (see remark on section 3.2).

The stronger positive trends found in the OSIRIS datasets are consistent with the

positive relative drifts of these datasets with respect to the MIM (Section 3.2). This suggests that differences in relative drifts may contribute to the larger trend estimates.

30. Caption Fig. 13: all error bars are full lines, not dashed or dotted on my printout.

See reply to comment 10.

31. line 418: how are the five layers above the tropopause defined?

This was indeed confusing and we changed the wording and sentences in the manuscript as follows:
“The ozone profile datasets generally start at an altitude of 10 km (Table 1) which can be an issue when the tropopause is lower than 10 km, particularly at higher latitudes. For the stratospheric column calculations, data were screened out if the three lowest 1 km layers above the tropopause had missing data. In combination with the larger uncertainties of limb measurements below 20 km, this selection criterion causes larger spreads in the profile and stratospheric column trends. “
At higher latitudes, some ozone profile measurements contain missing values above the tropopause. To account for this issue, we applied a screening criterion based on the missing 1 km layers above the tropopause. Stratospheric columns were only calculated when at least 3 km above the tropopause was valid ozone data. Missing data inbetween two valid data points (if any at all) was handled by the merged datasets themselves. No interpolation was done or needed in this analysis.
Because missing values in these high-ozone layers could potentially affect the resulting column values and the derived ozone trends, we additionally compared our approach with the method described by Sofieva et al. (2026). Here we first calculate the ozone trends of the profile ozone datasets and then integrate the columns. The comparison of both layer and column ozone trends is presented in Figure A6. The results show that the two methods produce very similar trends, particularly for the Full Regression analysis, indicating that the treatment of missing values has only a minor impact on the overall trend estimates. Therefore, results obtained with both methods are provided.

32. line 473 ends with two full stops.

Removed the second full stop.

33. Although the paper is predominantly based on satellite ozone retrievals, I somewhat miss the link with trends calculated from ground-based datasets. For the total ozone trends, you could point out if and where the satellite-retrieved total ozone trends differ from the WOUDC trends. And for the stratospheric (profile) ozone trends, you might compare those with the trend estimates appearing in Jonas et al., ACP, 2026 (https://doi.org/10.5194/acp-26-8089-2026) and Mirallie et al., egusphere, 2026 (https://doi.org/10.5194/egusphere-2026-113).

This is indeed important and we included some additional text for the comparison with WOUDC and the other total ozone dataset, as well as the stratospheric ozone trends

from Mirallie et al. (2026). We included a section in the discussion about the Mirallie et al. paper, but we didn't include the Jonas et al. paper, due to its focus on the Arctic which we do not consider.
"A comparison of our results, using the LOTUS setup (Fig. A7) with recent ground-based studies (Mirallie et al., 2026) shows a general agreement in the vertical structure of the ozone trends. In both studies, the largest positive trends are found in the upper stratosphere, consistent with ongoing ozone recovery. Mirallie et al. (2026) show clear negative trends between 20 to 30 km in the tropics and more mixed results in the northern hemisphere. This mostly mirrors our results although a slightly different proxy setup was used. The most notable difference is a slightly positive trend in the northern hemisphere above 20 km in our results, while Mirallie et al. show mixed results depending on the station (Figs. 20-23). This might highlight the need to look more closely at longitudinally resolved trends in a future study."

Replies to Referee 3:

34. In general the work is very carefully carried out and well explained. The use of dynamic proxies in the MLR model can be argued for or against but in my opinion is a good choice if the aim is to identify a trend due to EESC. Results such as those shown in figure 7 and figure 12 illustrate the significance of this choice. I am slightly surprised you don't highlight more that southern hemisphere mid latitude total ozone has a significantly positive trend from about 15-40 degrees even without dynamic proxies, this seems like a very important finding to me.

    This is now mentioned in the manuscript text. See comment 18 for the specific changes.

35. I have only one serious concern, which is the use of the three different SAOD proxies. I must say that I am not at all convinced that this is valid.
    The purpose of this proxy should be to quantitatively account for the *chemical* effect of extra aerosol surface area in the stratosphere. The dynamical effects should be captured by your dynamical proxies. GLOSSAC is latitudinally resolved so there should be no issue there. Figure 6 panels (i) (j) and (k) seem completely implausible to me. If the SAOD proxies are intended to capture both chemical and dynamical consequences of particular eruptions in a totally unspecified way then there is a large danger of introducing "fudge factors" that don't have a clear physical meaning. The approach taken here seems to me to somewhat defeat the purpose of MLR.
    Therefore I would suggest either clearing up the roles of these proxies in much greater detail (by which I mean the physical explanation and not just the statistical effect on trends), or keeping just one SAOD proxy for all periods based on GLOASSAC, or deleting periods from the record that MLR doesn't work properly for.

    We clearly indicate that the SAOD proxy does not stand alone for the chemical effect due to sulphate aerosol, but also covers the dynamical and radiative response from the enhanced aerosols (Line 273). For a given aerosol loading, the dynamical response can be highly variable and this is the main reason to use different SAOD proxies. Another reason would be that the aerosol surface area density (SAD) is responsible for the ozone depletion rather than the extinction coefficient and the relationship of SAD to SAOD might be different for different eruptions/conditions. We added some additional text to the paper describing, why we use three separate proxies for our regression.
    Line 269 (track changes version):
    “**Volcanic eruptions have different impacts on stratospheric circulation, chemistry, and ozone distribution, depending on magnitude, season, and latitude of the event (Stolarski et al., 2006; Chehade et al., 2014; Weber et al., 2018, 2022). This variable response is illustrated in panels i-k of Fig. 6. The Australian wildfire (end of 2019) and HTHH had a stronger initial impact in the SH than in the NH because the volcanic aerosols (and stratospheric water vapour) were transported mostly into the SH resulting in an ozone loss (band of negative values in panel i). The Pinatubo eruption had a stronger impact in the NH (boreal winter) as illustrated by the band of negative coefficients from the tropics to the polar region (panel j, Aquila et al. (2014)). The dynamic response to the El Chichón eruption showed the opposite effect. Due to**

changes in the SH circulation the dynamical ozone increase exceeded the chemical ozone loss caused by the eruption (see panel k, Zerefos et al. (1994); Angell (1997)). However, nearly all El Chichon coefficients are not statistically significant at 95% confidence level.."

36. It is very important to note that the APARC Hunga report found only quite modest chemical effect on ozone.
Østerstrøm, F. F., M. L. Santee et al. (2025): Effects of the Hunga eruption on stratospheric ozone and related trace gases. In APARC, 2025: The Hunga Eruption Atmospheric Impacts Report [Yunqian Zhu, Graham Mann, Paul A. Newman, William Randel (Eds.)]. APARC Report No. 11, WCRP Report No. 10/2025, DOI: 10.34734/FZJ2025-05242, available at https://aparc-climate.org/publications/aparc-report-no-1
The finding that tropospheric trends are not making a significant contribution to total ozone trends is also very important.

The chapter of Oesterstrom and Santee (2025) is already listed. In the discussion section the following is stated in the manuscript (Line 482 in the track change PDF): "The HTHH-SAOD proxy accounts to some extent also for the effect from the enhanced water vapor following the underwater eruption of HTHH (Vömel et al., 2022; Millán et al., 2022). Changes in transport/circulation had a larger influence on ozone than altered chemistry due to enhanced chlorine activation following the eruption (Oesterstroem and Santee, 2025)."

37. Line 21 It's not quite right to say "ie, the ozone hole", because you were talking about both hemispheres.

We slightly changed the sentences to remove this incorrectness.

38. Figure 1 **–** The caption should state what the colors are marking.

The colors indicate the viewing geometry of observation of each instrument. This was added to the caption of the figure.

39. Figure 1 **–** As a comment, the different datasets are not very independent with heavy reliance on SAGE II in the first part of the record. This would influence the regression results because the fit is over the whole period.

This is really important, the datasets are not independent in the first period at all. This reflects the limited availability of high-quality global ozone profile measurements during that period. However, despite using the same instrument data, the merged datasets differ in their retrieval versions, bias corrections, merging methodologies, sampling, and quality control procedures.

40. Lines 103-105 The satellite total ozone products have been extensively validated against

each other and against ground based data so they are not very independent.

This is true and we removed "independent" from the sentence.

41. Figure 2 - I really like this figure which sets things out very clearly, but the caption should state what the colors are marking.

The caption was extended, indicating that the colors highlight different instrument families.

42. Figure 2 – There is a minor inconsistency in the labels of OMPS-NP, where NOAA SBUV specifies Suomi-NPP and NOAA-20 but SBUV NASA doesn't.

Fixed

43. Line 108 "we need all ozone profiles in units" – please re-word slightly – it's more of a choice you've made.

We changed the sentence as follows:
In this study, all ozone profile datasets are expressed in units of number densities [molec cm^{-3}] as a function of geometrical altitude [km].

44. Lines 121- 140 It's great that you looked into this effect, which shows care and attention to detail.

Thank you!

45. Lines 139-140 You should include a short comment about what would be the general cause of why this might happen, is it departures from the zonal mean?

We changed the text to better address the comment, as follows:
"The significance of these trends decreases with altitude, showing the most significant trends at 25 km. This is possibly because of stronger vertical ozone or temperature gradients near this altitude, which increase the uncertainties from the non-linear ozone unit conversion and vertical axis transformation."

46. Lines 158-159 Again, I think a short comment is needed here about what would cause such effects. Does the fact the blue line in Figure 5 is slightly higher than the red line suggest that the more satellite instruments are included, the differences become greater?

We agree that an increasing number of contributing satellite instruments could potentially lead to larger differences between merged datasets due to differences in long-term stability. However, the comparison shown in Fig. 5 does not indicate a clear

dependence on the number of instruments alone, because the 1996 to 2024 line covers a longer time period than the 2000 to 2024, which includes more instruments (Fig. 2), but the 1996 to 2024 line shows lower drift uncertainties.

47. Line 165 You should include here a reference to Table 2 so the reader can check what each of the abbreviated proxy names refer to.

We added a references for Table 2.

48. Figure 6 comments:

1. I very much like this figure which I wish appeared in all papers which use MLR, in order to show that the regression results appear physically plausible and prevent the process turning into a 'black box'.
2. You need to say in the caption that the proxies have been normalized to 0-1 etc

   We added the information about the normalization to the caption:
   All proxies were normalized to -0.5 to 0.5 apart from the SAOD proxies (i), (j) and (k), which were normalized to 0 to 1.

3. What is the horizontal bar next to each proxy name representing?

   This is where SWOOSH does not have enough data in the high altitudes and the MLR does not work well anymore.

4. The solar response looks weak given how prominent the solar cycle is in long-term global ozone? (eg Figure 1 of Weber et al. 2022)

   We compared our coefficients with Weber et al. 2022 Fig. 5. We see that in Weber et al. the QBO response is stronger than the solar response by a factor of two The same relationship we see in our results. We see the maximum response from our solar proxy at approximately same latitudes as in Weber et al. 2022.

5. What is the grid cell in the tropical UTLS that shows up in panels a, b, c, g, h and j? Is it real?

   Those grid cells are at around 9 km in the tropics and with that below the tropopause. Because of this, we don’t think they are real, but rather result from fewer observations as well as cells in which the MLR does not work well anymore. We added the tropopause in this figure to highlight this.

6. A comparison of panels i, j and k show very little resemblance. Why is the effect of the aerosol so different?

   The influence of SAOD from volcanic activity heavily depends on the strength,

latitude and time of the year of the eruption. For this reason, we see very different results for the three different aerosol proxies and decided to split the SAOD into three separate SAODs. See also comments earlier.

49. Line 278 But wouldn't the long-lasting perturbations be caused by the aerosol remaining at elevated levels for longer, which would be captured by GLOSSAC? If this is not so then do you need to add a lag term?

It's not clear to which part of the manuscript this question has been raised.
A time lag was not used in recent studies Sofieva et al. (2026), Godin-Beekman et al. (2022), Weber et al. (2022). We use a latitudinal dependent SAOD, which accounts for the delay in the aerosol loading with latitudes due to the meridional circulation. looking at our time series we do not see any indications for a presence of a time lag, but it might be something that needs a more in-depth analysis (beyond the scope here).

50. Line 270 What do you mean by "absolute" here?

The sentence was changed so it only says "ozone timeseries", as a reminder that we use absolute units and not anomalies in our analysis.

51. Line 283 Is the fact that the ozone datasets agree well shown anywhere?

This is not shown in the paper. Weber et al. 2025 showed that there is good agreement between the satellite data after bias correction. It is only shown exemplary for the stratospheric columns (after bias correction) in Fig. 11 of our paper.

52. Lines 285-289 I find this discussion a bit confusing because the orange and blue lines agree very well in the SH from about 15 degrees to 35 degrees, and only then start to diverge, but in the NH they are different at 30 degrees, which is the opposite of the text. (It sounds like the text is referring only to the higher latitudes).

We added "40 poleward°" to the text to indicate which part of the southern hemisphere we are talking about. In the northern hemisphere at 30° we indeed see a larger difference between both methods, but this is still well within the uncertainties.

53. Lines 283-289 Also in this discussion, I would have thought that the fact the SH trends are significantly positive at many latitudes would be worth mentioning?

We now mention the significant trends in the southern hemisphere in comparison to the non-significant trends for the second method:
"The trends in the SH 40° (poleward) increase on average to +3 to +4 DU per decade (+1.3 to +1.5% per decade) when using the full proxy setup, while also being significant. With the LOTUS settings, SH trends are lower remaining at about +1.5 DU

per decade and insignificant above 45°S (see also Fig. A2 for the trends in % per decade).”

54. Line 291 "The more positive trend … in their and our … " – more positive than what?

We changed the text for more clarity:
“The full regression using additional dynamical proxies and three SAOD proxies shows above 40°S a stronger positive trend than the LOTUS regression (without dynamical proxies and a single SAOD proxy)”

55. Lines 308-312 I'm not clear what point you're making here. It is true that the LOTUS proxies were not designed for the poles, but as you say, they make a big difference in high mid-latitudes.

This part describes the trend model of the LOTUS regression, which does not work well in the polar regions compared to the Full Regression Model. We show that with the additional dynamical and the three SAOD proxies, the model works very well at high and mid-latitudes as well as in the polar regions. To improve on the statement, we changed the manuscript as follows:
“It is imperative to note, that the LOTUS setup (panel b) was developed for extra-polar latitudes and thus is not suited for polar regions, which is why it only serves as an illustration here.”

56. Line 314 Some minor re-wording is need here, because the "early phase" sounds like it might mean the early history of the ozone hole in the 1980s.

We added some text to refine the sentence:
“Polar ozone recovery was first identified in the early phase of the ozone hole season (September)”

57. Line 321 The influence of what, sorry?

We added: “influence of the dynamical proxies” to the sentence.

58. Lines 323-325 The effect on the ozone hole was very modest according to the APARC report.

This is mentioned in the discussion section

59. Lines 328-330 I can't see that you have demonstrated this, perhaps you could make this more explicit.

This is shown in Fig. 8 by the huge changes in ozone trends, when splitting the SAOD

into three separate proxies. Our reply to comment 1. explains it in more detail.

60. Figure 8 The arrangement of the panels isn't very user-friendly. Panels a-b-c-d are natural, but not d-e-f which would be easier to compare if they were in the same row or column.

That is correct. We tried various setups for this figure and even split up the figure into two, but this way it was not clear that panel d) was the same in the next figure. This is why we decided to include the SAOD split into Fig. 8.

61. Figure 8 Comparing panels a and c, by eye, the extra proxies seem to improve the fit in some years (2012, 2021) but don't help much in others (1982, 2008, 2018, 2019). Can you comment on that?

We don’t expect the dynamical and split AOD proxies to catch every minima and maxima of the time series, but rather the most prominent ones. The years 2018 and 2019 seem to have improved but not perfectly.

62. Figure 8 There is a huge difference in the quality of the fit between (b) and (d), but not so much between (d), (e) and (f). Why does the use of the third proxy make the recovery trend so much greater? Are you sure this is not an artefact?

The important distinction here is that the third AOD proxy is the split between the Pinatubo and the HTHH period. The HTHH period shows much lower values in comparison to the Pinatubo period, splitting these periods and then normalizing them improves the fit for the HTHH proxy for the 2000 to 2024 time period by not being overshadowed by the Pinatubo period. The larger post-ODS peak trends seen in panel d (+9%/decade) is in close agreement with the result from Weber et al. (2022), 12%/decade, indicating that the years 2020-2022 are an episodic excursion from the long-term ozone recovery due to the ODS decline.

63. Lines 359-363 Why would this be?

We have not examined the smaller SH positive trends of GOZCARDS in detail, but we did reproduce the results of Sofieva et al. 2026, using deseasonalized ozone data. We think this difference comes from the seasonality that was either removed by the deseasonalization or not caught by the MLR.

64. Lines 370-371 Wouldn't the use of the climatological value mean your trends will be affected by a rising tropopause height over this period, and/or tropical expansion into higher latitudes? Do you allow for different seasons?

Since we use quite broad latitudes with 10° steps and the changes of the tropopause are

rather small, we don't think that the use of the climatology will influence the calculation of the stratospheric column strongly. But we want to look into this in more details when using datasets with a better resolution (beyond of scope here).

65. Line 376 What is "bias corrected" here?

We added a short description of the bias-correction after the sentence:
"The stratospheric ozone column was bias-corrected by subtracting the mean during the 1998–2008 period from each dataset (anomaly) and adding back the multi-instrument mean to obtain bias-corrected absolute ozone values for each dataset."

66. **Figure 12 – It is very encouraging to see southern hemisphere stratospheric ozone is recovering so strongly in the IUP regression, although equally concerning that this is not true in the NH. Can you comment on this?**

This is most probably due to the asymmetrical ozone trends in the two hemispheres already evident in Fig. 9. Previous studies have shown that hemispheric differences in stratospheric circulation lead to an asymmetry in stratospheric HCl trends, with a faster decline in the Southern Hemisphere (Chrysanthou et al., 2025; Dubé et al., 2025). The associated decline in inorganic chlorine reduces ozone loss and therefore leads to a stronger ozone recovery in the SH. This aspect is already discussed in the manuscript.