Author's final response

Total ozone trends at three northern high-latitude stations

L. Bernet, T. Svendby, G. Hansen, Y. Orsolini, A. Dahlback, F. Goutail, A. Pazmiño, B. Petkov, A. Kylling

Dear referees, dear editor,

thank you for the constructive and elaborated comments and suggestions to our manuscript. We have taken your remarks and comments into account. With these changes, we think that the manuscript has significantly improved. The referee's comments are given in blue italic typeface, our responses are given in black, and the corresponding text in the manuscript in grey, with changes marked in bold.

Kind regards,
Leonie Bernet (on behalf of all co-authors)

Contents

1 Author's response to referee #1

1.1 General comments

> *The study of Bernet et al. (2022) provides total ozone trends at three high latitude stations for the recent period 2000-2020. While a start of the ozone recovery has been recently observed at mid-latitudes and Antarctic (e.g., WMO2018, Godin-Beekmann et al., 2022), it was still not the case for Northern high-latitudes. In the recent study of Weber et aL (2022), zonal mean 60-90°N total ozone trend in March from satellite measurements is still observed non-significant. It is therefore highly scientifically relevant to give an up-to-date ground-based perspective on ozone trend in the Arctic. Furthermore, the use of a multiple regression model to explain as much as possible the ozone variability makes also the paper well in scope with the ACP objectives. The paper is also very clear and well structured.*
> *I therefore recommend the publication of this study in ACP, after some comments and questions are addressed (mainly on the combined data sets / drifts with satellites / predictors).*

Thank you very much for your positive feedback and your constructive comments. We have followed your suggestions to improve the manuscript accordingly.

1.2 Specific comments

Section 2.4: Combined ground-based total ozone data

> *It's a good idea to combine the ground-based data sets to have a more complete temporal coverage. But I have a few comments/questions:*
> *Before combining, it would be good to show that the different ground-based measurements do not show significant bias among them, inter-comparing data in temporal coincidences.*

We added a comparison between GI and DS data (new Fig. A2, see comment Line 82 by Referee 2). Comparisons between GUV-DS and GUV-SAOZ have been presented by Svendby et al. (2021), who found no systematic biases. We added this information to the manuscript in the beginning of Sect. 2.4.
(see also comment l. 118 by Referee 2)

GUV data have been validated against DS and SAOZ data by Svendby et al. (2021), who found no significant biases between the datasets. GI data is evaluated in Appendix A.

> *You use one technique as a baseline then successively fill the temporal gaps by other techniques. But why not using all measurements, also when different techniques are measuring at the same time, and then make a daily mean? Of course, in the way it is now (SAOZ at sunset/sunrise, the other instruments at +/-2h around local noon), the instruments are not co-located in time to make the average. But as you use anyway SAOZ in the combined time-series together with the +/-2h around noon averages, without any correction, I guess it would be valuable to make the "total" daily mean averages? If there is a reason to choose your approach of filling gaps instead of taking everything available, could you add an explanation?*

For each station, we chose an intsrument as the "main" instrument (Brewer at Oslo and Andøya, SAOZ at Ny-Ålesund). Only for days with missing data, measurements from other techniques are used. These other techniques (GUV and GI) are assumed to be less reliable, and are therefore only used when required.

We thus prefer not taking "total" daily mean averages, but use the other techniques as complementary information. We added this information to the manuscript.
The co-location in time is only an issue in Ny-Ålesund (where SAOZ is used as "main" instrument). Brewer and GUV instruments are most reliable for smallest solar zenith angles (SZA) and we therefore use Brewer and GUV measurements around noon, even though SAOZ measures only at sunrise and sunset. However, the diurnal cycle in total ozone is assumed to be small in Ny-Ålesund, because daily variation in total ozone is mainly determined by the emission of precursors (Antón et al., 2010), which is assumed to be negligible in remote places as Ny-Ålesund.

Brewer DS and SAOZ have been used in multiple studies (e.g. Goutail et al. (2005), Hendrick et al. (2011) and Scarnato et al. (2010)) and are therefore used as baseline in GBcomb. Measurements gaps are then filled with GI and GUV data.

Section 3: Time series comparisons

> *Drifts after 2015/Fig.4: Because the drifts/jumps between GBcomb and the satellites are different depending on the stations, and similar at each station looking at different satellites, it looks like the drifts are in the GBcomb products (for Andoya and Ny-Alesund), which could be an issue for interpreting the trends obtained in Sect.5 (especially for Ny- Alesund). It could be worth to evaluate quantitatively the drifts, and discussed them with the trends.*

We added some more discussion on the drifts to the manuscript. However, it is difficult to decide which data are "right". When comparing Brewer with GUV data at Andøya, no clear drift or anomaly is visible. However, a similar drift is visible when comparing GUV with ERA5 data. This suggests that it is not only related to the Brewer-SL changes, but that there are probably some drifts also in ERA5 and/or satellite data. We added this to the text.
Also in Ny-Alesund we cannot observe any clear drift in one of the ground-based instruments. Looking at Fig. 4(i) (or Fig. 3(i) in the revised manuscript), it could also be related to larger differences in the two extreme years 2016 and 2020, with an overestimation of ozone by the satellites (compared to GBcomb) in 2016 and an underestimation in 2020, which gives the impression of an apparent drift. Nevertheless, the ground-based data (GBcomb) may slightly overestimate ozone at Ny-Alesund in 2020 due to high Brewer DS measurements. This may be related to some technical issues in 2019 at Ny-Alesund and the missing calibration afterwards, due to Covid-19 related travel restrictions. We added this to the manuscript.

Sect. 3: In Andøya we observe a drift compared to some satellites from 2014 to 2016, **which may be related to the issues with the Brewer standard lamp during this period (Sect. A4). However, we found no drift when comparing Brewer with GUV data, which makes the attribution of the origin of the drift difficult. Furthermore, we observe a similar drift when comparing ERA5 with GUV data, suggesting that ERA5 is drifting in this period of new assimilated satellite products (as mentioned above).** The drift at Andøya occurs mainly compared to GOME2 and ERA5 and is less visible compared to OMI and SBUV. In Ny-Ålesund, we observe a drift in opposite direction starting in 2016. Further analyses would be required to investigate these differences. Interestingly, most satellites **overestimate ozone at Ny-Ålesund in 2016 and** underestimate ozone in 2020 compared to the ground-based data, **which both were years with extreme cold stratospheric conditions. However, GBcomb might slightly overestimate ozone in 2020 due to high Brewer DS values after 2019. The Brewer-instrument at Ny-Ålesund was last calibrated by IOS in 2018 and a new calibration and inspection of the instrument will hopefully reveal potential problems.**

> *In Fig. 2, comparisons between Brewer and ERA-5 at Andoya show a kind of "jump" between 2015-2019, which is related in the discussion and figure to some disturbance observed in SL measurements (that are corrected, but apparently not perfectly since the jump is still observed). Could it be that this impacts the comparisons with satellites as well (Fig.4h)? In Fig.4h, GBcomb are used, but the Brewer measurements are 84% of the GBcomp for Andoya. Maybe the comparisons between Brewer and GUV (suggested in the comment above on Sect. 2.4), can confirm if the drif/jump is also observed with local measurements (and not due to e.g. grid resolution of ERA and Satellites)? Is there a way to adjust the correction that has been made in order to remove the drift/jump at Andoya? For Ny-Alesund: how do the SL measurements look like there?*

We agree that it is not clear whether the SL correction was able to completely remove the anomaly in Brewer data at Andøya. We adapted the text accordingly (sect. A4) and mention it also in the discussion on satellite comparison (see previous comment). However, similar anomalies are observed when comparing GUV with ERA5 data at Andøya. This suggests, that ERA5 can also be responsible for those differences in that period.
In Ny-Ålesund, we only use the well established DS measurement method and no GI data, and did therefore not investigate in detail the SL data. Please also note, that Brewer in Ny-Ålesund is only used starting in 2013 and constitutes only 20% of the combined time series in Ny-Ålesund.

Sect. A4: Such irregularities can **partly** be handled thanks to the regular calibrations and the SL correction in the GI retrieval **(see Appendix A4), but we observe larger differences to ERA5 in this period (Fig. A3). However, similar anomalies are observed when comparing ERA5 to GUV data at Andøya, suggesting that ERA5 is showing a drift from 2015 onwards.**

> *However, when looking at anomalies Fig4e-f, it looks like the trends obtained from the satellites would not be so different than the ones from GBcomb. Maybe it would be interesting to calculate those satellite trends (e.g from 2005) at the location of the stations and see if they agree with GBcomp's trend starting in 2005?*

We agree that it would be interesting to investigate satellite trends at the stations. However, the satellites do not cover the whole time series (except the merged SBUV) and investigating satellite drift correction and/or merging issues to derive corrected satellite trends would be beyond the focus of our study.

> *It looks like SBUV and Era-5 show a similar drift in 2000-2005 at all sites, suggesting that the drift is due to SBUV and Era-5 and not to the GBcomp in that period. Maybe it is worth saying in Sec. 2. 6 which satellites are assimilated in Era-5? Is it mainly SBUV in this period /latitudes?*

SBUV is indeed assimilated in that period in ERA5, but also EARTH PROBE TOMS and ERS 2 GOME until 2003 and ENVISAT SCHIAMACHY starting in 2003 (see Figure 7 in Hersbach et al. (2020)). The drift seems to stop when Aura data is assimilated starting in the end of 2004. We added a sentence to the manuscript.

Furthermore, ERA5 and SBUV seem to drift from 2000 to 2005 compared to the ground-based data at all stations. Afterwards, this drift is not visible anymore, suggesting that the drift may be corrected with the beginning assimiliation of Aura observations in ERA5 (Hersbach et al., 2020).

Section 4: Multiple linear regression

> *Error covariance matrix: you give higher uncertainty to monthly means that show higher standard deviation of the measurements, so with higher variability within the month. Could this also lead to less weight given for events with higher variability (e.g. ozone loss due to VPSC would give a high ozone variability within the month), therefore this could lead to minimizing the impact of proxies (which is not desired)? How this error covariance matrix impacts the obtained R2, trends,...?*

That's true, we assume that variable months are more uncertain. However, we think that this choice is appropriate. Taking your example of higher variability due to VPSC-related ozone loss, the variability would be high if it is not the whole month that shows depleted ozone values. The monthly mean would thus not be representative for the whole month. We therefore think that using a higher uncertainty for such monthly means is appropriate.

Concerning the impact of the error covariance matrix on the trend, please refer to Bernet et al. (2021). They showed with an artificial time series that using appropriate uncertainties in the error covariance matrix during anomalous periods improves the trend fit (Table 1 and Fig. 4 therein).

> *Tested predictors: a local proxy that has been proven to have significant influence on the ozone variability at high latitudes stations is the equivalent latitude (See Vigouroux et al., 2015, Fig.4). It can take into account the O3 short term variability due to the fact that the station is in/out of the polar vortex. Did you try this (or alternatively the potential vorticity)?*

We agree that equivalent latitude would be an interesting local predictor and we considered investigating it as additional proxy. However, due to limited time, we were not able to include equivalent latitude in the final analysis. We hope that we can include it in future work.

> *Final choice of predictors: it is decided to keep using T50 although it correlates with TropP (0.51), and to not use VPSC because it correlates to EHF (-0.33). It looks quite arbitrary to have these opposite decisions. The use of T50 despite the correlation is motivated by the increased R2 (from 0.91 to 0.96 at Oslo). It is said that the use of VPSC improved the fit residuals at Ny-Alesund, but the R2 improvement is not given: how much would the VPSC inclusion increase R2 at Ny-Alesund? If negligible, then it should be fine to remove it from the model. If not, then the correlation motivation does not seem in line with the choice made for the T50 inclusion.*

As you mentioned, including VPSC reduces the fit residuals in some years at Ny-Ålesund. The R^2_{adj} of annual trends, however, is not improved when including VPSC. We added this information to the text.

T50 improves R^2_{adj} at all stations and almost all months. When using VPSC, the only month with a slight improvement in R^2_{adj} is March in Ny-Ålesund. However, VPSC in Ny-Ålesund correlates significantly not only with EHF, but also with TropP and T50. For those reasons, we decided to keep T50 and to exclude VPSC as a predictor.

Finally, we do not include VPSC because of the weak but significant correlation to EHF (Fig. 5) **and no general improvement in R^2_{adj}**, even though we observe that the fit at Ny-Ålesund **is improved (smaller residuals)** in some extreme years when VPSC is included (not shown).

> *L-265-267: comment: I find interesting and consolidating your results that the same behavior is observed with FTIR total ozone (Vigouroux et al., 2015): EHF significant at the polar sites, but not at mid-latitudes. Solar, QBO, and ENSO also insignificant or very small at polar sites.*

Thank you for this remark, we added this information to the manuscript.

This confirms results by Bahramvash Shams et al. (2019) based on ozonesonde measurements in the Arctic **and by (Vigouroux et al., 2015) based on FTIR measurements.**

> *6 and 7: the solar cycle parameter is found negative. To my knowledge the impact of solar cycle on total ozone is expected to be positive (solar maximum = increase of O3; see e.g Weber et al. 2022,...). This is not what is found at Oslo for annual trend, and even more for March trend. I don't understand the discussion on February trend at Oslo (l.313-319): the explanation seems to say that O3 observed maximum coincide with the maximum of solar cycle, but the parameter is found negative (Fig.7). Could it be that the MLR gives a wrong interpretation of solar cycle influence? How would be the February trend without the solar cycle included? Maybe more realistic (now it looks like an outlier in Fig.9). The proxy with few cycles included (less than 2 for 11-years solar cycle 2000-2020), can be "by chance" interpreted to have an influence while "physically" they don't. It was also the case for short time-series in Vigouroux et al. 2015. There is no "scientific reason" that the impact of solar cycle would be so stronger in February, is there? So it would be good to give the trend obtained without this solar cycle. Especially, because the time-series starts at a solar maximum and ends at a minimum, the inclusion or not of this proxy will modify the final trends (if the parameter is large, as it is the case in Feb in Oslo).*

We agree that the discussion about the solar cycle contribution to February ozone in Oslo is a bit confusing. Further analyses with longer time series and a larger geographical coverage would be required, which would be beyond the scope of this study. We therefore decided to modifiy and/or remove these sentences and added a note concerning further required investigations.

~~Furthermore, we observe that the periods with less variable February values seem to coincide with the solar maxima. The rather stable February years in the beginning of the time series (2000-2003) coincide with the first maximum of the solar cycle, and higher ozone occurs again during the second solar maximum in the years 2013 and 2015. This may explain the strong contribution of the solar predictor to ozone in February at Oslo as visible in Fig. 7.~~ **Furthermore, we observe a significant negative contribution of the solar predictor to February ozone at Oslo as visible in Fig. 7.** Longer time series **and additional stations** would need to be investigated to analyse this effect further.

> *Lower R2 at Andoya and Ny-Alesund (l.282-284): could you try with VPSC and equivalent latitude included? From FTIR stations (Vigouroux et al. 2015), the R2 was larger at polar sites (including Ny-Alesund) compared to mid-latitudes because of this larger variability (less dominated by noise in agreement with what you discussed l.279-280). When the variability is well explained, R2 is more easily higher.*

When VPSC is included, the R2 in March at Ny-Ålesund is slightly increased, but almost no effect is visible in other months (see also previous comment ("Final choice of predictors")). It would be interesting to investigate in future studies whether including equivalent latitude would improve R2 at polar sites.

Section 5: Trend results

> *5.1: You obtain slightly increasing trends with latitude. Could you give the obtained trends with SBUV (only satellite you use covering the 2000-2020 period) at the 3 sites to check if this is also observed by the satellite? Because of the observed drifts in Fig.4, it might be good to consolidate your results.*

As mentioned earlier (see comment to Section 3), we think that it is beyond the frame of our study to investigate satellite trends (which would include merging and drift issues).

> *5.2: February in Oslo: see comments above on the solar cycle impact.*

See comment above.

> *5.2: October positive trends: note that zonal means trends around 80°N in October were still found negative in Morgenstern et al. (2021). So interesting indeed that you don't find this over Scandinavia. It would be interesting to see what is observed at other longitudes. You could add this reference to motivate your work on regional trend.*

Thank you for this remark, we added the remark and the reference to the text.

In contrast to our results, Morgenstern et al. (2021) found significant negative trends in September and October at 80° N, using satellite zonal means. These contrasting results suggest that it is important to investigate regional trends and not only zonal means.

1.3 Minor or technical comments

Abstract

> *- l. 10-11: specify "positive annual trends at Andoya,..."; and give uncertainties*

Done.

> *- l.11: "no significant annual trend at Oslo (0.1%/dec + uncertainty)*

Done.

Introduction

> *- l.30-31: I guess drifts can also be observed in ground-based measurements.*

We agree, but ground-based observations have the advantage that they can continuously be recalibrated and have therefore not the typical instrumental degradation or drifts as satellites have. We replaced "drifts" by "degradation".

> *- l. 48: “in the best possible way’. Maybe too assertive. Can be moderated by “... and define a state-of-the-art set of predictors that explains the natural ozone variability at the three stations”; or something similar.*

Done, thank you for the suggestion.

Section 2

> *the GI method is applied to Oslo and Andoya. Why not to Ny-Ålesund ?*

Unfortunately, the GI measurements have not been configured yet at Ny-Ålesund. We hope that this can be done in future.

> *you give the available months for GUV and SAOZ; maybe give them also in Sect. 2.1.1 (DS) and 2.1.2 (GI); or summarize all this in a Table instead, as you wish?*

We added a table to summarize the availability of all months and adapted the text accordingly.

> *you give the uncertainties of SAOZ measurements (l.112): give also the uncertainties of DS, GI, GUV (It could be in the same Table as for period of measurements suggested in the previous comment?)*

We added the uncertainties for DS and GUV in the text. GI uncertainties are discussed in the appendix.

DS: **The precision of DS measurements is 0.15 % (Scarnato et al., 2010) but the accuracy relies on regular calibrations (Svendby et al., 2021).**
GUV: **GUV ozone retrievals depend on accurate calibrations and Svendby et al. (2021) showed that the standard deviation between Brewer and GUV in Oslo and at Andøya are within 2.5% for the period 1995 to 2019.**

Section 4

> *259: “full” trend: I guess you mean “annual trend”? I would use annual, it is more common (same for title 5.1,...)*

We used the term "full trend" to distinguish clearly from annual mean trends, but we agree that the term "annual trend" is more common and adapted the whole manuscript accordingly.

> *To my opinion, it is easier for the reader to include Fig B1 and B2 in Fig. 7.*

We agree that including Fig. B1 and B2 in Fig. 7 could be helpful. However, we decided to only show the results for Oslo as an examplary case for the predictor selection. With this we avoid an overload of figures. Showing predictor results for the stations Andøya and Ny-Ålesund here would imply to include those stations also in Figs. 5 and 6, which would lead to too many figures. We prefer showing the Oslo results as an example to give an insight into the trend predictor selection and therefore prefer keeping Figs. B1 and B2 in the Appendix.

Section 5

> *1: provide the uncertainties on the trends also in the text.*

Done.

> *296-298: I would give the numbers in the two reported satellite studies to better compare with your results.*

We agree and added the trend numbers from the two reported studies.

For example, Sofieva et al. (2021) found positive trends over Scandinavia **of 1 to 5 % per decade depending on altitude** based on merged satellite ozone profile data from 2003 to 2018 (their Fig. 10), and Coldewey-Egbers et al. (2022) reported significant positive total ozone trends **of 1.2% per decade** in the North Atlantic sector based on merged total ozone data from 1997 to 2020 (their **Table 3** and Fig. 3).

> *299-302: I would give the trends and uncertainties from Svendby et al., to help the reader to see quickly the improvements on uncertainties obtained using the MLR.*

Done, thank you for the suggetsion.

Furthermore, Svendby et al. (2021) analysed total ozone trends with a simple linear regression from 1999 to 2019, using ground-based GUV measurements at the same three stations as in our study. They found similar **trend magnitudes**, but slightly larger uncertainties **(1.5±1.8% per decade at Oslo, 0.5±2.6% per decade at Andøya, 1.2±2.4% per decade at Ny-Ålesund).**

Conclusions

> *Typo l.358: "In conclusion,"*

Done.

> *358: "the urgency": too strong to my opinion. Maybe "the need"?*

Done.

2 Author's response to referee #2

> *This is the first review of the manuscript "Total ozone trends at three northern high- latitude stations" submitted by Bernet et al. to the ACPD. The paper is focused on the assessment of regional ozone trends as measured by ground- based instruments in the Arctic. Multiple publications state that the Arctic environment is rapidly changing due to the impacts of climate change. The changes in the global circulation and stratospheric temperatures have been reported in the literature, however, the full extent of the processes contributing to regional ozone variability in the Arctic is not well established. Moreover, the impact of the halogens on the spring-time stratospheric ozone depletion is intermittent in the Arctic due to the typically high instability of the polar vortex that influences the formation of the Polar Stratospheric Clouds that must last into the spring (or transported to the lower latitudes) when sunlight can activate the rapid ozone destruction mechanism. The intensive ozone depletion events are still hard to predict; thus, the analyses of the past records are of great importance. Several papers published in ACP in 2022 found that total column ozone trends derived from the combined satellite records are not the same across the Arctic. Therefore, it is important to compare these results with the trends from the long-term ground-based records that are known for their stability and regular quality assurance of observations. The authors present the study of long-term records from three stations in Norway that collect observations in one section of the Arctic. They also represent different latitudes that might be impacted by different processes, including being inside or outside the polar vortex and downwind of different pollution sources. The attribution of the long-term changes in the ozone column is performed statistically using multiple proxies that represent chemical and dynamical processes that impact stratospheric ozone. A thorough evaluation of the correlations between proxies and their respective ability to explain observed variability is needed to determine statistically significant trends. The annual and monthly trends are derived to address the regional and seasonal variability in total column ozone in the Arctic and to uncover the signals of the ozone recovery under the Montreal protocol guidance. The paper is well written, the figures are used well to illustrate the main points of the discussion. However, there may be the need to have additional figures in the Supplemental material since there are several occasions when the analyses are mentioned but the figures or summary of the results (i.e. in the table format) are not available. The published literature is properly referenced. This paper can be published after the authors address the following comments.*

Thank you very much for the recommendation and the constructive criticism. We address your comments in the following.

2.1 General comments

> *1) It would be important to discuss the location of the stations with respect to the vortex position before and after it breaks. Are stations typically within or outside of the vortex during spring season observations?*

All the three stations can be influenced by the polar vortex, depending on the year. We added a comment on the vortex location for the different stations in Section 2.

Ozone at all three stations is influenced by the polar vortex in winter and spring, with varying degrees depending on the year.

> *2) I did not find any discussion about the potential impact of the long-range pollution transport to the Arctic with respect to the total ozone variability. Is there any evidence for tropospheric ozone increases in the Arctic?*

Currently, Arctic air pollution, including precursors for tropospheric ozone production, is still dominated by anthropogenic emissions transported from Europe, Asia and North America (Marelle et al., 2018). These emissions show decreasing trends especially in the 1990s (e.g. Sharma et al. (2013)). In future, however, anthropogenic activity in the Arctic may increase due to the decline in sea ice, because Arctic shipping may then become a major source of ozone pollution in summer (Marelle et al., 2018). The consequences for total ozone have still to be investigated.
Tropospheric ozone in the Arctic is still low (e.g. Gaudel et al. (2018)) compared to other areas in the world. We therefore neglect tropospheric ozone variability in our trend analysis. We added a sentence about this to the manuscript.

Our study concentrates on stratospheric ozone. The contribution of tropospheric ozone to total ozone is assumed to be small in the study time period and locations (see Gaudel et al. (2018) and Sharma et al. (2013)) and tropospheric ozone variability is therefore neglected in our trend analysis.

> *3) It would be of interest to the reader to have information about the magnitude of the trends in the proxies, or at least if the trends are positive or negative.*

We agree that it would be interesting to investigate possible trends in the predictors. However, such trends would not affect our results, because the effect of the predictors is in fact excluded from the signal in the regression. The final trend is the remaining ozone trend that can not be explained by the predictors (see l. 182 before modifications). Investigating all predictors for trends would thus not affect the study's outcome and would be beyond the scope of the present study.

2.2 Technical comments

> *Line 79. The authors refer to the SL abnormalities being "corrected therein". What does it mean? Are data corrected daily, monthly, or yearly? Did the authors have to correct the data for this paper? How large were the corrections and if any were applied with a step change? Please explain.*

The SL-correction refers to the calibration that is continuously performed using SL-measurements in order to check the stability of the instrument. This is a standard procedure for Brewer instruments. We reformulate the sentences to make that clear.

All DS measurements are regularly calibrated with **daily** standard lamp (SL) measurements. For the SL-correction, the intensities of an internal halogen lamp (SL) are measured at the same five wavelengths as for the ozone measurements. The SL produces a stable and continuous light spectrum. Any variations visible in the SL-measurements would also affect the ozone measurements and are therefore corrected in **the routine DS calculation**.

> *Line 82 (and Appendix A). It is not clear if the GI method is applied to the Brewer observations. I am guessing it is as the discussion is under section 2.1 Brewers.*

Yes, the GI method is applied to Brewer measurements, we adapted the text to make that clear.

The **Brewer** DS method is limited to clear-sky conditions and solar zenith angles (SZA) below 72°. Therefore, we use the global irradiance (GI) method to retrieve ozone **from Brewer measurements** in cloudy conditions and/or for larger SZA at Oslo and Andøya.

> *Line 82 (and Appendix A) continued. Was the GI processing applied to the Brewer data for the first time and you are presenting these new data in this paper? If this is true, please also spend more time discussing the validation of the Brewer GI data (it can be done in the appendix/supplemental material). I would move most of the discussion of the GI data (including Figure 2) to the Appendix. Also, please include plots for other stations (Oslo?) to demonstrate the stability of the GI method in comparison to the Brewer DS measurements. Otherwise, please provide the reference to the paper that describes the application of the GI method for the Brewer data processing and demonstrates its validation.*

The Brewer-GI data is presented in a peer-reviewed manuscript for the first time, but the data have been used in various scientific reports for many years. The GI processing method is similar to the method used for GUV data by Svendby et al. (2021).
For further validation of GI data, we added a figure in which GI is compared with coincident DS data.
As suggested, we moved all discussion of the GI data (including Fig. 2) to the Appendix. We also added the GI-comparison-figures for Oslo. We adapted the text accordingly.

Fig. A2 shows the comparison of GI daily means with coincident DS data at Oslo and Andøya. On average, we observe an absolute difference between GI and DS of around 1% at both stations, indicating a good agreement. We observe a slightly higher difference before 2005 at Oslo and after 2015 at Andøya. These episodes coincide with rapid changes in the SL values.

> *Line 85. Does the reference to the "Ground-based UV radiometers" include both Brewers and GUV?*

Svendby et al. (2021) use the method to which we refer here for GUV radiometers and not for Brewer GI measurements. We added "GUV" in the text to make that clear.

The same method has recently been used to derive total ozone from Ground-based UV **(GUV)** radiometers by Svendby et al. (2021).

> *Line 87. Figure 2 comments. Please add a plot of the difference between ZS and GI data and reduce the y axes range to see the details. To make it more clear, the data could be averaged into monthly means to reduce the daily noise.*

We adapted the figure to show only smoothed data (with a 30days moving mean window) and removed also the SL data in Fig. A2 (the SL data is now only shown in the new figure that compares GI vs. DS (Fig. A2), see your comment l. 118). With these changes, the difference between GI and ZS is well visible and we don't think that a seperate ZS-GI comparison is necessary.

> *Line 90. Please quantify "slightly" by providing %bias for ZS/ERA5 comparisons for the reader's reference. In my opinion, the ERA 5 should not be used as truth to evaluate the quality of observations, but it can help to investigate sudden changes in the record that can be associated with the instrumental artifacts (or the changes in the assimilated data).*

We added a value to quantify the difference between ZS and ERA5.
We agree that ERA5 can itself show drifts, but it can help to investigate potential changes in the ground-based record. Nevertheless, we assume that some of the difference between GI and ERA5 at Andøya from 2015 to 2018 is also due to ERA5, and added this to the manuscript (see also comments by Referee 1 about Section 3).

... with ERA5 than ZS measurements, especially at Andøya **(3.3% difference between ZS and ERA5)**.
... **we observe larger differences to ERA5 in this period (Fig. A3). However, similar anomalies are observed when comparing ERA5 to GUV data at Andøya, suggesting that ERA5 is also showing a drift from 2015 onwards.**

> *Lines 91-92. Please provide information about the cause of the degradation of the instrument*

The exact reason for the problem with Brewer at Andøya is not known, despite several inspections by IOS. A jammed screw was detected in 2015, which caused problems for movement of the Brewer micrometer, but this could probably not explain the steep SL drift after 2015.

> *Line 118. Are secondary data corrected for biases from the primary data? Was a comparison between the primary and secondary records done for each station during the overlap periods? Alternatively, the references for the publications describing the comparisons can be provided.*

(see also first comment by Referee 1)
We added a comparison between GI and DS data (new Fig. A2, see your previous comment (Line 82)). Comparisons between GUV-DS and GUV-SAOZ have been presented by Svendby et al. (2021), who found no systematic biases. We added this information to the manuscript in the beginning of Sect. 2.4.

GUV data have been validated against DS and SAOZ data by Svendby et al. (2021), who found no significant biases between the datasets. GI data is evaluated in Appendix A.

> *Line 124. How do sunset and sunrise data compare to the noon observations when all three are available? Was the standard deviation of the mean used to screen the data?*

We compared SAOZ data with noon-measurements from Brewer and GUV that were available at the same dates, and didn't find any systematic difference when comparing noontime Brewer data to sunrise and sunset SAOZ data. We assume the diurnal cycle of total ozone in the Arctic to be small. Antón et al. (2010) showed that diurnal variability in total column ozone is largely determined by variations in surface emissions of precursors, which is negligible in remote areas such as Ny-Ålesund.
Yes, the standard deviation of the daily mean was used to screen the data and to reject outliers, as described in the same section (l. 125, version before modifications).

> *Line 117. What are the dates for change in satellite record assimilation over the period of observations?*

Various satellite products have starting and ending dates during our study period. An overview of all assimilated ozone observations used in ERA5 is given in Hersbach et al. (2020) (Figure 7 therein).

> *Line 144. Please rewrite the sentence (i.e. "seasonal cycle in ground-based observations").*

Done.

The seasonal cycle and the interannual variability is clearly visible, as well as the measurement season of the ground-based observations with missing GBcomb data in winter months at Andøya and Ny-Ålesund.

> *Line 150. Was there a change in 2014 in the selection of datasets assimilated into ERA5?*

There was indeed a change of assimilated ozone observations in 2014: the beginning of METOP-B GOME-2 ozone observations, and the end of NOAA 16 SBUV-2 ozone observations (Hersbach et al., 2020). We added this information to the manuscript.

However, we observe that ERA5 reports more total ozone than the other datasets at all stations starting in 2014, **which coincides with a change in assimilated satellite data in ERA5, as indicated by Hersbach et al. (2020) (new assimilation of METOP-B GOME-2 and change in assimilated SBUV-2 data).**

> *Line 157. What is the distance between satellite overpasses and Ny-Alesund station location in early spring?*

At Ny-Ålesund, the overpass distances are approximately around 65km for GOME2a, around 180km for SBUV and 25km for OMI. Distances are larger close to winter months (in February, March and October). We don't think that the overpass distance have an effect on the observed drifts.

> *Line 182. Should there be a reference to the previous publications that discussed the meaning of using detrended vs. not detrended proxies for trend analyses?*

A good explanation of the advantage of not detrending the predictor time series is given by Weber et al. (2022) (end of Section 3 therein). They state that detrending the predictors would add all trends to the linear trend term, which would make the attribution of ODS-related changes impossible. We added the reference to the sentence in our manuscript.

By including the predictors in the regression without detrending them, any trend that is due to long-term changes in one of the predictors is removed from the ozone time series. The remaining, unexplained trend **can then be attributed** to changes in ODSs **(Weber et al., 2022)**.

> *Line 250. March makes sense for the co-linearity issue as this is the time for the vortex breakup and mixing of different air masses. However, results for September are not clear (unless there are tropical air intrusions to the high latitudes). Please add a discussion here about the processes that the T50 tracer represents. Plotting residual for a model fit when including/excluding T50 proxy can potentially help with identifying*

years when it improves the fit.

We compared model residuals with and without the use of T50. From a simple investigation it looks like the years with the largest residual improvements are often years with special stratospheric activities such as major stratospheric warmings. Confirming this claim would require more in-depth analyses, but it seems to be consistent with the objective of using T50 as predictor to represent large-scale stratospheric circulation patterns.
It is more difficult to understand the reason for the correlation between T50/TropP and other predictors in September. For this, further analyses using model and backtracking data would be required, which is beyond the scope of our study.

Line 256. Please add an explanation of why you think it is "improved". How much improvement did you find (use the R2 or explained variability) Also, the authors decided to keep the T50 proxy in the regression based on the improvement to one station fit. What is the "physical process" that underlines the preference for including T50 vs VPSC proxy? The temperature and ozone are highly correlated in the stratosphere, and T50 has a seasonal cycle. Does T50 inclusion in the model reduce the seasonal terms?

We refer to the improvement of the residuals in specific months when the VPSC-predictor is considered. We adapted the sentence to make this clear.
Concerning T50, we decided to keep it because it improves R2adj at all stations and in most months (see also comment of Referee 1 "Final choice of predictors"). We added this to the text.
T50 represents not only chemical activity, but also large scale circulation changes in the stratosphere. We added a sentence to the manuscript.
Please note that we use deseasonalized T50 data, as mentioned in Table 1.

l. 251: First, we decided to use T50, even though it is correlated to TropP, because including T50 substantially improves the adjusted coefficient of determination (R^2_{adj}) of the **annual** model fit (e.g. from 0.91 to 0.96 at Oslo) **and for most months**.
l. 261: Finally, we do not include VPSC because of the weak but significant correlation to EHF (Fig. 5), even though we observe that the fit **is improved (smaller residuals)** in some extreme years when VPSC is included at Ny-Ålesund (not shown). **The temperature-dependent chemical activity is already covered by using T50, which in addition represents circulation changes in the stratosphere.**

Lines 260-264. Would other proxies have a larger contribution to the fit if T50 is not included?

Yes, some of the variability that is accounted for by T50 would be captured by other predictors when T50 is not included.

Line 276. Would adding the "rejected" proxies improve the model fit for the summer months? The cumulative EFH is constant for May-August. What other proxy represents the dynamical variability of ozone in the summer months? Was the contribution of tropospheric ozone variability considered for explaining the increased noise in total column ozone in summer? Would the model fit improve if the seasonally (three or 4 months) averaged data are used instead of monthly records?

We observe indeed a slight improvement of R2 by approximately 0.1 at Olso and Andøya in June and July when the AO and NAO predictors are included. However, due to the existing correlations, including those

predictors would lead to multicolinearity issues in many months. As discussed in the manuscript, we therefore prefer not including them.
We did not investigate tropospheric ozone, as described in the reply to your previous comment (general comments 2)).
When investigating seasonal trends, the R2 did not improve compared to the monthly trend fits, and was even worse in winter months.

> *Line 280. Please explain what processes might be contributing to the noise.*

Noise is referring here to normal natural variability, we added this to the text.

Furthermore, the interannual variability is generally low in summer and may be dominated by **natural variability or** noise, as suggested by Brunner et al. (2006).

> *Line 283. I cannot see the higher variability in the provided plots. Please add information about the standard deviation to the plots or the Table.*

We are referring here to larger anomalies in Andøya and Ny-Ålesund, as stated in l.146 (version before modifications). We further checked the standard deviation of the GBcomb data, which is also larger for Andøya and Ny-Ålesund. We added the standard deviation to the first row in Fig. 4 and added a sentence in Sect. 3.

The standard deviation of GBcomb data increases with latitude of the station.

> *Line 291. Instead of using "good" please provide the SE of the detrended data in %. This can help the reader with an understanding of the improvements in the model performance.*

We added the standard error expressed as percentage of the mean ozone value for each station and adapted the text and figure texts (Fig. 7 and Fig.9 in the revised version) accordingly. Further, we corrected a typo in the SE calculation which leads to slightly larger values (no significant changes).

The standard errors of the residuals (SE_{res}) of **1.53** DU (Oslo) to **2.63** DU (Ny-Ålesund) **indicate that the predicted values differ from the true ozone values by less than 1%.**

> *Line 305. I recall that the LOTUS model allows using of the entire dataset to retrieve the seasonal trends by adding the seasonal components to the proxies. This approach can reduce the uncertainties in the derived trends by including information from other seasons. Was this technique considered for this paper? Was the measurement error covariance matrix used to analyze the observed records?*

Deriving the seasonal trends directly using the LOTUS method was not possible in our case due to the measurement gaps in our time series. When using this option in the LOTUS model, seasonality is added to the linear term, which is "trying" to fit seasonality to the data. This is not working well with our data due to regularly missing data in winter months. We therefore computed seasonal trends by investigating time series of each month separately.

> *Lines 305-314. There seems to be a repeat of the information. Please consider rewriting to outline the results for all stations based on one season at a time. It might be useful to keep in mind the processes that affect ozone variability in spring vs. other seasons.*

Thank you for this input. We agree that this part is repetitive and modified the whole paragraph. We grouped the descriptions by season.

> *Line 318-319. The authors report that the solar cycle (SC) proxy is significant for model fit in February only. Some papers (i.e. Labitzke and Loon, 1988) found positive correlations between polar stratospheric temperatures or geopotential height data and SC when QBO was in its western phase, and the opposite was found for the eastern phase of the QBO. So, some of the SC attributions to ozone variability are probably represented by other proxies (i.e. T50 and TP) that you are using for model fit. Still, the contribution to ozone variability in February only is hard to justify because the same process should affect ozone variability in other months. Could the method of combining records for February in Oslo be impacting the time series? What is the percentage of DS vs GI observation is used in February for the combined record? Doe this percentage change from year to year? Another possibility is that T50 or TP proxies are not representative of atmospheric variability in February in Oslo. Or there is a sampling bias of ozone records in February.*

Thank you for these interesting comments and ideas. We agree that the SC contribution in February would require much more investigations, including longer time series and a larger geographical coverage, which would be beyond the scope of this study. We therefore decided to modifiy and/or remove these sentences and added a note concerning further required investigations.

~~Furthermore, we observe that the periods with less variable February values seem to coincide with the solar maxima. The rather stable February years in the beginning of the time series (2000-2003) coincide with the first maximum of the solar cycle, and higher ozone occurs again during the second solar maximum in the years 2013 and 2015. This may explain the strong contribution of the solar predictor to ozone in February at Oslo as visible in Fig. 7.~~ **Furthermore, we observe a significant negative contribution of the solar predictor to February ozone at Oslo as visible in Fig. 7.** Longer time series **and additional stations** would need to be investigated to analyse this effect further.

> *Line 324 Please re-write this sentence (they mean “polar regions”?). Please provide the reference to the publications that discussed pre-2000 trends in the Arctic region.*

We rewrote the sentence and added some reference.

Ozone trends in spring months are of special interest in polar regions, ~~as they~~ **because those regions** experienced strongest ozone depletion in the pre-2000 phase (e.g. Solomon, 1999).

> *Lines 330-332. Yes, it is remarkable that 3 stations have very similar trends in March. Was the PSC proxy tested to improve the trend model fit for analyzing data in March in an attempt to reduce the uncertainty? In section 4.3 it is not clear if the selection of final proxies was done based on the model fit to the entire dataset (an annual record) or if it was also assessed by using the fit to the monthly records.*

Yes, we tested the monthly trends with and without the use of the VPSC-proxy. The predictor selection was mainly based on the annual record, but the monthly predictor contribution was also considered. The only month with a slight (but not significant) reduction of uncertainties when using VPSC was March in Ny-Ålesund. In other months, using VPSC had almost no effects, but led to an overfit of the model. Furthermore, VPSC in Ny-Ålesund correlates significantly with EHF, TropP and T50, and shows a significant correlation with EHF also at Oslo, as mentioned in Sect. 4.2 (l. 255 before modifications). For those reasons we decided not to include the VPSC predictor in our final model (see also comment of Referee 1 "Final choice of predictors").

Finally, we do not include VPSC because of the weak but significant correlation to EHF (Fig. 5) **and no general improvement in R$^2_{adj}$**, even though we observe that the fit at Ny-Ålesund **is improved (smaller residuals)** in some extreme years when VPSC is included (not shown).

> *Line 352. Please define "high" (for example, better than 0.9).*

We agree, all annual trends have values > 0.95, which we added to the text.

with high coefficients of determination (R^2_{adj} $>$0.95)

> *Line 365. Please replace the NDACC link with www.ndacc.org The "demo" link will be going away soon, but the "ndacc" link will be preserved for a long time.*

Done, thank you for this comment.

2.3 Comments for Figures

> *Figure 8. The trend results are provided in the Figure in % and DU, but Standard Errors (SD_res) are provided only in DU while the plots are using % in y-axes. It might also help the reader if all results can be also summarized in the Table.*

We added SD_res in percent (see previous comment, l.291) and added a table summarizing all trend results.

> *Appendix A. General Comment. Thank you for using the color scheme in the plots that are friendly for vision-impaired people.*

We use in all qualitative figures the "vibrant" qualitative colour scheme provided by Paul Tol, which is colourblind safe. More information about this colour scheme is available at `https://personal.sron.nl/~pault/`.

> *I would also consider reducing the range of X-axes in all panels (except for T50) in Figures 8, B1 and B2. Some boxes are impossible to discern. The range for the T50 panel can be different and the note can be added to the Figure caption.*

Thank you for this idea, we adapted the range of the x-axes in Fig. 7 and added a note to the figure caption.

Fig. 7 caption: **Note that the range of the x-axis is different for T50 than for the other predictors.**

> *A2. Since GI data are corrected for the SZA bias, it is possible that Brewer data also has a bias (i.e .in Oslo where Brewer is a single spectrometer). The change in the data processing is noted from the 4-wavelength to one wavelength ratio method at large SZAs. However, there is still a need to calibrate data by using the DS measurement. What is the range of the SZAs for the GI/DS comparisons?*

The upper SZA level for the DS/GI comparison is 76.2 degrees. In Svendby et al (2021) it was only used Brewer DS data with ozone slant column below 1100 DU where the effect of stray light was negligible. However, there were few days with values above 1100 DU and the GUV SZA correction was almost the same whether we applied a Brewer slant column limit or not. We cannot know for sure that the GI correction for SZA above 76.2 degrees in the winter is correct, but comparisons to our GUVs and satellite data indicate that the agreement is good for all seasons.

> *A3. It might be useful to show a plot (for both stations) where the GI/DS total ozone ratio is plotted as a function of the SZA before and after each correction (SZA and cloud correction).*

We added a figure with the DS/GI ratios as a function of SZA and CLT that were used to derive the linear coefficients (new Fig. A1).

3 References

Antón, M., M. López, A. Serrano, M. Bañón and J. A. García (2010). “Diurnal variability of total ozone column over Madrid (Spain)”. In: *Atmospheric Environment* 44.24, pp. 2793–2798. DOI: 10.1016/j.atmosenv.2010.05.004.

Bernet, Leonie, Ian Boyd, Gerald Nedoluha, Richard Querel, Daan Swart and Klemens Hocke (2021). “Validation and trend analysis of stratospheric ozone data from ground-based observations at Lauder, New Zealand”. In: *Remote Sens.* 13.1, pp. 1–15. DOI: 10.3390/rs13010109.

Gaudel, A., O. R. Cooper, G. Ancellet, B. Barret, A. Boynard, J. P. Burrows, C. Clerbaux, P. F. Coheur, J. Cuesta, E. Cuevas, S. Doniki, G. Dufour, F. Ebojie, G. Foret, O. Garcia, M. J. Granados-Muñoz, J. W. Hannigan, F. Hase, B. Hassler, G. Huang, D. Hurtmans, D. Jaffe, N. Jones, P. Kalabokas, B. Kerridge, S. Kulawik, B. Latter, T. Leblanc, E. Le Flochmoën, W. Lin, J. Liu, X. Liu, E. Mahieu, A. McClure-Begley, J. L. Neu, M. Osman, M. Palm, H. Petetin, I. Petropavlovskikh, R. Querel, N. Rahpoe, A. Rozanov, M. G. Schultz, J. Schwab, R. Siddans, D. Smale, M. Steinbacher, H. Tanimoto, D. W. Tarasick, V. Thouret, A. M. Thompson, T. Trickl, E. Weatherhead, C. Wespes, H. M. Worden, C. Vigouroux, X. Xu, G. Zeng and J. Ziemke (2018). “Tropospheric Ozone Assessment Report: Present-day distribution and trends of tropospheric ozone relevant to climate and global atmospheric chemistry model evaluation”. In: *Elementa* 6. DOI: 10.1525/elementa.291.

Goutail, Florence, J. P. Pommereau, F. Lefèvre, M. Van Roozendael, S. B. Andersen, B. A. Kåstad Høiskar, V. Dorokhov, E. Kyrö, M. P. Chipperfield and W. Feng (2005). “Early unusual ozone loss during the Arctic winter 2002/2003 compared to other winters”. In: *Atmospheric Chem. Phys.* 5.3, pp. 665–677. DOI: 10.5194/acp-5-665-2005.

Hendrick, F., J. P. Pommereau, F. Goutail, R. D. Evans, D. Ionov, A. Pazmino, E. Kyró, G. Held, P. Eriksen, V. Dorokhov, M. Gil and M. Van Roozendael (2011). “NDACC/SAOZ UV-visible total ozone measurements: Improved retrieval and comparison with correlative ground-based and satellite observations”. In: *Atmospheric Chem. Phys.* 11.12, pp. 5975–5995. DOI: 10.5194/acp-11-5975-2011.

Hersbach, Hans, Bill Bell, Paul Berrisford, Shoji Hirahara, András Horányi, Joaquín Muñoz-Sabater, Julien Nicolas, Carole Peubey, Raluca Radu, Dinand Schepers, Adrian Simmons, Cornel Soci, Saleh Abdalla, Xavier Abellan, Gianpaolo Balsamo, Peter Bechtold, Gionata Biavati, Jean Bidlot, Massimo Bonavita, Giovanna Chiara, Per Dahlgren, Dick Dee, Michail Diamantakis, Rossana Dragani, Johannes Flemming, Richard Forbes, Manuel Fuentes, Alan Geer, Leo Haimberger, Sean Healy, Robin J. Hogan, Elías Hólm, Marta Janisková, Sarah Keeley, Patrick Laloyaux, Philippe Lopez, Cristina Lupu, Gabor Radnoti, Patricia Rosnay, Iryna Rozum, Freja Vamborg, Sebastien Villaume and Jean-Noël Thépaut (2020). “The ERA5 global reanalysis”. In: *Q. J. R. Meteorol. Soc.* 146.730, pp. 1999–2049. DOI: 10.1002/qj.3803.

Marelle, L., J.-C. Raut, K. S. Law and O. Duclaux (2018). “Current and Future Arctic Aerosols and Ozone From Remote Emissions and Emerging Local Sources—Modeled Source Contributions and Radiative Effects”. In: *J. Geophys. Res. Atmospheres* 123.22, pp. 12, 942–12, 963. DOI: 10.1029/2018JD028863.

Morgenstern, Olaf, Stacey M. Frith, Gregory E. Bodeker, Vitali Fioletov and Ronald J. van der A (2021). “Reevaluation of Total-Column Ozone Trends and of the Effective Radiative Forcing of Ozone-Depleting Substances”. In: *Geophys. Res. Lett.* 48.21, e2021GL095376. DOI: 10.1029/2021GL095376.

Scarnato, B., J. Staehelin, R. Stübi and H. Schill (2010). “Long-term total ozone observations at Arosa (Switzerland) with Dobson and Brewer instruments (1988–2007)”. In: *J. Geophys. Res. Atmospheres* 115.D13. DOI: 10.1029/2009JD011908.

Sharma, S., M. Ishizawa, D. Chan, D. Lavoué, E. Andrews, K. Eleftheriadis and S. Maksyutov (2013). "16-year simulation of Arctic black carbon: Transport, source contribution, and sensitivity analysis on deposition". In: *J. Geophys. Res. Atmospheres* 118.2, pp. 943–964. DOI: 10.1029/2012JD017774.

Solomon, S (1999). "Stratospheric ozone depletion: A review of concepts and history". In: *Rev. Geophys.* 37.3, pp. 275–316. DOI: 10.1029/1999RG900008.

Svendby, Tove M., Bjørn Johnsen, Arve Kylling, Arne Dahlback, Germar Bernhard, Georg Hansen, Boyan Petkov and Vito Vitale (2021). "GUV long-term measurements of total ozone column and effective cloud transmittance at three Norwegian sites". In: *Atmospheric Chem. Phys.*, pp. 1–29. DOI: 10.5194/acp-21-7881-2021.

Weber, Mark, Carlo Arosio, Melanie Coldewey-Egbers, Vitali E. Fioletov, Stacey M. Frith, Jeannette D. Wild, Kleareti Tourpali, John P. Burrows and Diego Loyola (2022). "Global total ozone recovery trends attributed to ozone-depleting substance (ODS) changes derived from five merged ozone datasets". In: *Atmos. Chem. Phys.* 22.10, pp. 6843–6859. DOI: 10.5194/acp-22-6843-2022.