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.

Here are several comments:

General comment:
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?

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?

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.

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.

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. 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.

Line 85. Does the reference to the “Ground-based UV radiometers” include both Brewers and GUV?

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.

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).

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

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.

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?

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

Line 144. Please rewrite the sentence (i.e. “seasonal cycle in ground-based observations”).

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

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

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?

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.

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? Lines 260-264. Would other proxies have a larger contribution to the fit if 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? Line 280. Please explain what processes might be contributing to the noise. 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. 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. 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? 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. 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. 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. 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. Line 352. Please define “high” (for example, better than 0.9). 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. 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. Appendix A. General Comment. Thank you for using the color scheme in the plots that are friendly for
vision-impaired people.
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.
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 SZA.
However, there is still a need to calibrate data by using the DS measurement. What is the range of the SZA for the GI/DS comparisons?
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).