Friday, 27th March 2026

Dear referee,

We thank you very much for your thorough review of our work. Please find below our answers to your comments, embedded in your review with a different color for eased readability. Lines references to the paper are for the new manuscript version and not the diff version.

All the best,

Caroline Jonas on behalf of all co-authors.

Review of Jonas et al. *"Detection of ozone recovery in the Arctic from ground-based measurements"*

General comments

The submitted manuscript reports indications of ozone recovery in the Arctic, based on a carefully considered use of available ground-based instruments at multiple locations.

Firstly, the data records from the ground-based stations are compared to long-term satellite records to identify any outliers.

Secondly the individual stations are joined together in appropriate regional composites using analysis of ozone spatial variability at different heights in the atmosphere.

Thirdly, 2000-2024 trends are calculated using a multilinear regression model with selected proxy terms.

In the concluding section the results are compared in reasonable detail to other published work.

The topic is a very important one for atmospheric science and I have no hesitation in recommending publication in ACP once some comments below have been addressed. On the whole the work has been carried out very carefully and is well explained.

Specific comments

As the authors rightly point out, trend detection in the Arctic has proven very challenging due to the small signal compared to the variability. Secondly, over 2000-2024 we would expect to see significant circulation changes either forced or unforced (eg Arosio et al.) which have the potential to mask changes due to chemical depletion. Figure 13 (which I think is a very good figure) demonstrates the difficulty being faced because the trend is so small compared to the other proxy terms. Therefore it is hard for the reader not to wonder whether the trends are really just an artefact of the many proxy terms.

- The reason why we need to use proxies in the first place is precisely because the variability is large compared to the size of the trends. Accounting for variability by MLR with proxies is what enables to reduce the uncertainty of trends and to obtain significant trend results. This is particularly flagrant when observing the LOTUS trends in Figures C1 and C2: the LOTUS proxies are not sufficient to explain the large variability of ozone in the Arctic, and therefore, the trends uncertainties are very large and all the trends nonsignificant. The large variability of ozone in the Arctic is due to the specific chemical and dynamical conditions in this region (polar vortex, VPSC…). We have detailed in more length the use of each of our proxies in this context in the introduction and why they effectively explain this variability. The fact that we can obtain significant trends with reduced uncertainties, while we also see an improvement in the goodness of fit through better R^2 values (Figure 13) give us confidence that the trends are not an artefact and really represent the “recovery” of ozone, once other chemical and dynamical processes are accounted for. By the way, note that the fact that the R^2 term attributed to the trend is small in Figure 13 is not a bug but a feature: the trend term is not expected to explain the variability, so the size of this contribution only depends on the size of the trend, and we know that it is expected to be small.

As shown in the well-known "Weber plot" [eg Fig 4-13 in Chipperfield, Santee et al. 2018] interannual variability in Arctic ozone is largely controlled by dynamics which can be well represented by the accumulated eddy heat flux. Although the authors have gone to quite some effort to avoid over fitting, none the less they are left with multiple proxies that are essentially dynamic effects in different forms. The use of temperature as a proxy is slightly contentious because it isn't linked to any broader scale circulation features. I suspect here the temperature is capturing the zonally asymmetric features not picked up by the other proxies which are zonally symmetric.

- Concerning the use of stratospheric temperature as a proxy, it has been shown since the statistical analysis of Mäder et al., 2007 that temperature at 50hPa is one of the few essential proxies for explaining dynamical variability of ozone in the Arctic. This proxy was also used more recently in other Arctic ozone studies such as Bernet et al., 2023. It is true that the stratospheric temperature also impacts for instance the VPSC formation, but by analysing the signal of both proxies when they are both relevant, we find that they do not explain the same variability features in the time series. Moreover, detrending the temperature rather has the effect of reducing the zonal trend asymmetry or has no effect for the total column and the mid and upper stratospheric columns, so we do not think that the temperature is capturing purely zonally asymmetric features. We have emphasized the role and use of the temperature proxy in the introduction by referring also to the work of Mäder et al. 2007.

On the other hand, the comparisons of trends in different geographic areas, different seasons and different partial columns all add together to make a much more convincing story.

I would therefore request that the authors discuss these points somewhat more specifically than at present, in other words, why should the reader believe that these are real trends?

- See first comment above. We have also emphasized in the conclusion of the paper our confidence in our trends due to our analysis of the R^2 and trend uncertainties, as well as the coherence of the proxies throughout regions, layers and seasons.

A second point I would make is that it seems to me there are several unspoken assumptions in the approach used, which are all very reasonable but should be stated explicitly rather than just being implied. In this category I would include that individual stations with problems can be identified by comparing to the satellite dataset but that the drift in the satellite dataset can be identified with the combination of ground stations, and secondly, that monthly anomaly correlations between different stations imply the decadal trend is also correlated.

- We have made these assumptions more explicit in the paper where relevant, see below for each case in the technical comments.

Future work would be accounting in more detail for the seasonal and geographic differences.

The manuscript is on the whole very well written. In one or two places, according to my own personal taste the wording is slightly too informal for a published scientific paper.

Technical comments

Line 6 – I would suggest not using "drifted" as an adjective like this, instead say "drifts in ground-based data sets"

- done

Line 9 – I don't think you mean "natural", perhaps a better word would be "non-chemical"? Circulation changes influencing ozone resulting from GHG emissions aren't "natural".

- Natural variability here means variability due to natural processes (be they altered by climate change or not). This wording is used widely, e.g. in WMO 2022 chapter 4, section 4.1.1, so we have decided to keep it as is.

Lines 27-30 The discussion of negative trends in the lower stratosphere has generally been in the context of mid-latitudes though rather than polar regions. Van der Gathen et al 2021 is really a different topic.

- In these lines, we wanted to point that more negative trends in the lower stratosphere northern hemisphere have been attributed to two distinct effects: in the mid-latitudes (and indeed this has been the most discussed topic), it is due to the internal variability of ozone transport; in the northern polar region, it is due to the stratospheric cooling induced by climate change (that's where the von der Gathen reference comes in hands). We have tried to make our phrasing clearer in the new manuscript version, lines 29-32.

Line 37 (The recommendation for citing chapters of the ozone assessment is to list the two lead authors, eg Chipperfield & Santee et a;. 2022)

- done

Line 40 "Parallely" change to "In parallel".

- done

Line 41 "on it's own" change to "in its own right".

- done

Lines 57-59 As above, the "Weber plot" shows how well internnual variability in Arctic ozone can be represented by simply the accumulated Eddy Heat Flux. so accounting for the dynamics is absolutely essential for what you're trying to do. LOTUS perhaps is aimed differently.

- The first phase of LOTUS was focused on mid-latitudes trends and was therefore less insisting on dynamical processes. We have reformulated and extended all the introduction of the various proxies used in this work in a clearer manner. Dynamical proxies are indeed essential for our work. The Eddy Heat Flux is in fact used in our work as a proxy for the Brewer-Dobson Circulation, but a mistake occurred when reporting the proxy used for this process, we have written it was the Eliassen-Palm flux, although we ended up using the accumulated Eddy Heat Flux in the Northern Hemisphere as given by the OREGANO project database (https://www.iup.uni-bremen.de/OREGANO/proxydata/). This mistake has now been corrected.
  Note that those proxies are nevertheless strongly related. We would further like to point out as well that the strong correlation between the interannual ozone variability and the EHF in the Arctic can also be observed with the VPSC, as in Fig. 2b of Rex et al 2004. As explained in the paper in lines 474-482, when we remove the VPSC proxy, part of the variability explanation is taken up by the BDC proxy (EHF), although the total R^2 remains smaller than with VPSC.

Line 62 "trends uncertainties" replace with "the uncertainties of our trends"

- done

Line 80 Insert "and" before "6-hourly pressure"

- done

Lines 85-88 The description of NDACC seems a bit overblown, eg "a massive effort". You list the total number of stations and the record of 35 years but of course in your work you are only considering a selection of these stations and only for 25 years.

- We have shortened the NDACC description in this sense.

Table 1 expand the abbreviation "DOFS" in the caption

- done

Line 122 "lower" should be "below"

- done

Lines 148-151 This approach is ok but potentially is a source of problems for interpretation of the lower stratospheric partial column, compared to the alternative of using "tropopause-relative co-ordinates".

- Indeed, the fact that we use a fixed altitude level as top of the tropospheric column has an influence on what is calculated. In particular, the tropopause rising is a possible reason for the negative trend we observe. This is also not uncommon in tropospheric trends calculations (see Van Malderen et al 2025) This has been detailed in the text, both in lines 164-167 and later when we talk about the impact of detrending proxies, in particular the tropopause pressure, lines 507-508.
  Note that for the FTIR for instance, we cannot avoid mixing with the lower stratosphere because the profiles have a low resolution and the averaging kernels are not peaked enough to completely disentangle partial columns. This justifies our choice a partial column, to enable each to have about one DOFS.

Lines 159-160 This statement seems strange to me. Do you mean, different biases and starting dates would have an effect but you assume the effect is small? Please make this clearer.

- What we meant is that using anomalies instead of columns for the merging enables to merge regardless of possible biases between the datasets. However, the fact that different starting and finishing dates do not change results is indeed an assumption, as the anomalies can change depending on the period of reference with respect to which they are calculated. This is assumed to have a minor effect. We tried to make this clearer in the manuscript, see lines 177-178.

Lines 170-173 The "A" and "B" notation doesn't seem to be defined properly. You also use it in the plots (eg Figs 4, 5).

- We have tried to improve the notations for the different IASI products used, we hope it makes it all clearer, see lines 189-193.

Line 190 I like that you show the figure to help the reader more quickly understand your approach.

- Thanks!

Line 197 "spotting" would be better as "detecting"

Lines 196-199 As mentioned earlier this seems like an assumption to me, but a reasonable one.

- Indeed, we have made explicit that this is an assumption, see lines 221-225.

Figure 2 I like the idea of the figure but it doesn't seem very clear to me. I think the red dotted line is the bias and the solid horizontal line the zero-axis, but then where is the drift?

- We have updated the figure and the caption, hopefully making it clearer. The drift is the trend of the datasets plotted, which is the difference between the ground-based and satellite anomalies time series.

Lines 211-213 Is the different sampling frequency a problem in the troposphere, in that the ozonesondes are typically only make one flight a week?

- Indeed, this as been clarified in the text, see lines 241-242. We have numerically checked and justified in the text that the difference of time sampling doesn't change drifts by applying different filter on the minimal amount of sonde measurements per month. Note that this is not specific to the troposphere, however.

Figure 3 – I like the figure but would prefer it to be orientated vertically like the iconic NDACC figure. Also then you could mark the different height ranges but also the alternative ones too in a dotted line say.

- Thank you for this nice suggestion!

Lines 222-229 This finding is actually quite important for the ozonesonde community because it implies that more work is still needed (more than HEGIFTOM) to properly account for these changes.

- Yes indeed, this finding is important, and some co-authors of our paper are member of the sonde community and are aware of these issues. As of now, no direct solution was found to correct for these issues through reprocessing.

Line 223 "resp." change to "respectively"

- done

Line 235 – "recalibration" seems the wrong word

- done

Line 237 "below" would be better as "equatorward"

- done

Line 251 "problematic" is not quite right here, it would be better to say, e.g. "Having set aside the ground-based time series with identified problems"

- done

Lines 252-254 Could you explain the way you formed the "zonal band" means more clearly please? Do you mean just 60-90, what I would call the "polar cap"?

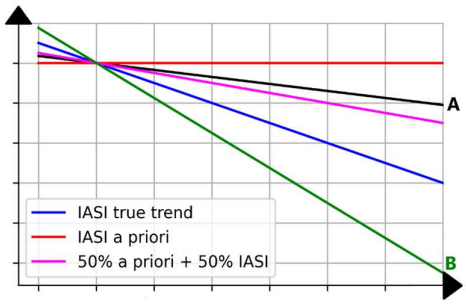
- The term “zonal band” is an abuse of language, we have replaced it by the phrase “zonal mean” everywhere in the text and defined what we mean by that, i.e., a weighted mean of all ground-based stations datasets within the polar cap in line 183.

Line 256 "is not drifted" replace with "does not show a drift"

Lines 257-258 Are you saying that it is well known that IASI has lost sensitivity over time in the high latitudes?

- No IASI sensitivity has not been degrading over time in high latitudes, as we have explicitly verified by computing trends of IASI DOFS in the troposphere. What we meant is that due to the low sensitivity of IASI, about 50% of the information in the troposphere originates from the IASI a priori which is a fixed a priori. Therefore, if there is a nonzero trend in the true tropospheric ozone, this will imply a “fake” drift when the sensitivity is low and a large part of the information comes from the a priori, as is the case in the troposphere. When the ground-based dataset is degraded to the IASI sensitivity through smoothing, the drift effectively compares the ground-based trend to the IASI true trend (blue line in the Figure below). Without smoothing, it is effectively compared to 50% of the

IASI a priori and 50% of the IASI (pink line below).



We previously assumed that this could in part explain the drift seen, but in fact the fixed IASI a priori actually hides a stronger IASI drift, as visible in Boynard et al 2025 (Table 3, 60-90 degree N). See lines 284-288.

Lines 259-261 Does this mean you shouldn't be doing what you are doing?

- No, we mean that the correct way to validate the satellite dataset is to account for the IASI sensitivity by smoothing. However, the smoothing formula cannot be applied on monthly mean as it would otherwise introduce additional errors. We have therefore not included it but we have added the effect seen in Boynard et al 2025, which would be an aggravation of the negative tropospheric drift by about 2%/decade. We have detailed all this in the manuscript, see lines 288-291.

Lines 269-271 I find this statement difficult to interpret. Do you mean that the drifts at individual locations all even out when added together? (When I first read the sentence I thought you meant you must have made a mistake).

- We assume that the individual satellite drift is the same at every location. Therefore, since the individual location drifts even out when merged and since the merged product drift has a very small uncertainty, it implies that the drift of the individual satellite on the whole arctic region considered is small. We clarified this sentence now in the manuscript, see lines 302-304.

Line 278 "ample" replace with "the many"

- done

Line 295 Should be "square kilometres"

- done

Line 296 Please re-word "Drowned out" – if they're truly "drowned out" you can't see them at all.

- done

Line 309 – 317 I think some additional discussion is required here. I think you are assuming that correlations in the monthly anomalies will imply the long-term trends are also equal? Is this reasonable?

- Indeed, this assumes that correlations are not only due to interannual variability. Since we are comparing the same variables (i.e., ozone anomalies) at different locations, we expect the same physical processes to cause the variability, and thus similar trends are expected for largely correlated locations. This assumption is confirmed by observing monthly anomalies timeseries that are highly correlated such as Kiruna and Andoya or Eureka and Alert for the total column. We have made this assumption more explicit in the text, see lines 338-341.

Line 361 "account for the natural variability of ozone" – really the proxies are "trying to account" because they are not totally successful, and rather than "natural" it's more the "non-chemical" variability of ozone.

- Same comment as before for the “natural”, but otherwise done

Lines 386-368 I assume you calculate the volume of PSCs across all of the Arctic, or do you mean only at the location of each station? This seems like a tricky point because you don't know whether the air has been exposed to PSCs earlier in the season.

- We calculate the global volume of PSCs in the Arctic. We have made this clearer in the text 9line 405-407) and in table 6. Indeed, this is a very broad and vague estimation that doesn’t consider the “history” of the air and its previous exposition to PSCs. Despite this, we find that the VPSC proxies has a very large contribution to the R^2 in many cases, which gives us confidence in the usefulness of this proxy.

Line 371 The use of temperature as a proxy seems unusual to me. (It has a major effect as shown in Fig13.) The problem with using it is that it doesn't give you any information about what is causing the temperature change, you're not linking it to broader processes.

- The stratospheric temperature was used as a proxy in previous studies of Arctic ozone trends such as Bernet et al. 2023, and in the statistical analysis of ozone proxies carried out in Mäder et al. 2007, it was found to be one of the most important proxies in the North polar region. It is true that it has a strong effect on the trends, which is the reason why we have included a discussion on the detrending of proxies. We have also added a more in-depth discussion of the use of proxies in the introduction, see lines 54-74.

Line 377 It would be worth mentioning here that the proxies are not de-trended and you will discuss this later

Line 380-382 Just a comment, this method was widely used in ozone trend studies some years ago but seems to have become less fashionable

- Yes but we think it is nonetheless totally trustworthy for our purpose

Lines 399-408 But what about if the tropopause has systematically risen over the last 20 years in some locations?

- This is reflected by the trend of the tropopause pressure and can be seen in Figures C1, C2, See added comment in lines 508-509.

Lines 409-410 This is an unusual thing to say – does it imply that you have chosen the wrong proxies? Could you have used more relevant proxies for this part?

- No, we have not chosen the wrong proxies in the case of this study as our main focus is on stratospheric ozone and we are primarily interested in the impact of Stratosphere-Troposphere exchange processes. It was therefore important to use a coherent set of proxies for the whole study. Furthermore, we find significant contributions to R^2 from e.g. the arctic oscillation, the temperature and equivalent latitude and even the solar cycle and tropopause pressure. Beside this, the tropospheric ozone variability is difficult to analyse and depends on many factors as chemical species changes (eg NOx decrease, CH4 increase, etc.), fire incidence or vegetation coverage change. Taking these into account requires to use a global simulation on long-term such as in Law et al. 2023 and lies beyond the scope of our paper, see lines 452-455. We also added a discussion of Law et al 2023 results in parallel of our results on the tropospheric trends, see lines 612-616.

Lines 441-448 I think this is a very good discussion. I feel some authors confuse themselves with what exactly they're trying to show.

- Thank you, we think this is very important too.

Figure 13 This is a very useful figure in my opinion.

- Thank you.

Lines 534-535 In theory the zonal asymmetry should be able to be captured if you were using the right proxies, shouldn't it?

- The upper stratosphere variability is not well explained by our model. However, by detrending the equivalent latitude (i.e., the polar vortex proxy), we find that the zonal asymmetry becomes stronger in Spring, which means that we are already explaining part of the asymmetry. This is included in the text now, see lines 591-592.