Atmospheric Chemistry and Physics Discussions

Atmos. Chem. Phys. Discuss., referee comment RC1
https://doi.org/10.5194/acp-2022-488-RC1, 2022

Comment on acp-2022-488

Corinne Vigouroux (Referee)

Referee comment on "Total ozone trends at three northern high-latitude stations" by Leonie Bernet et al., Atmos. Chem. Phys. Discuss., https://doi.org/10.5194/acp-2022-488-RC1, 2022

---

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).

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.
- 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?

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. In Fig. 2, comparisons between Brewen 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?

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?

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 SBUB in this

period /latitudes?

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,...?
- 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)?
- 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.
- 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.
- 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).
- 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.

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.

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

- 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.

3) Minor or technical comments:

Abstract:

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

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

Introduction:

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

- 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.

Section 2:

- the GI method is applied to Oslo and Andoya. Why not to Ny-Alesund ?
- 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?
- 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?)

Section 4:

- 259: "full" trend: I guess you mean "annual trend"? I would use annual, it is more common (same for title 5.1,…)
- To my opinion, it is easier for the reader to include Fig B1 and B2 in Fig. 7.

Section 5:

- 1: provide the uncertainties on the trends also in the text.
- 296-298: I would give the numbers in the two reported satellite studies to better compare with your results.
- 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.

Conclusions:

- Typo l.358: “In conclusion,”
- 358: “the urgency”: too strong to my opinion. Maybe “the need”?