<u>Review by Owen R. Cooper (TOAR Scientific Coordinator of the Community Special Issue)</u>

This review is by Owen Cooper, TOAR Scientific Coordinator of the TOAR-II Community Special Issue. I, or a member of the TOAR-II Steering Committee, will post comments on all papers submitted to the TOAR-II Community Special Issue, which is an inter-journal special issue accommodating submissions to six Copernicus journals: ACP (lead journal), AMT, GMD, ESSD, ASCMO and BG. The primary purpose of these reviews is to identify any discrepancies across the TOAR-II submissions, and to allow the author teams time to address the discrepancies. Additional comments may be included with the reviews. While O. Cooper and members of the TOAR-II Steering Committee may post open comments on papers submitted to the TOAR-II Community Special Issue, they are not involved with the decision to accept or reject a paper for publication, which is entirely handled by the journal's editorial team.

General Comments:
TOAR-II has produced two guidance documents to help authors develop their manuscripts so that results can be consistently compared across the wide range of studies that will be written for the TOAR- II Community Special Issue. Both guidance documents can be found on the TOAR-II webpage: https://igacproject.org/activities/TOAR/TOAR-II
The TOAR-II Community Special Issue Guidelines: In the spirit of collaboration and to allow TOAR-II findings to be directly comparable across publications, the TOAR-II Steering Committee has issued this set of guidelines regarding style, units, plotting scales, regional and tropospheric column comparisons, tropopause definitions and best statistical practices.
Guidance note on best statistical practices for TOAR analyses: The aim of this guidance note is to provide recommendations on best statistical practices and to ensure consistent communication of statistical analysis and associated uncertainty across TOAR publications. The scope includes approaches for reporting trends, a discussion of strengths and weaknesses of commonly used techniques, and calibrated language for the communication of uncertainty. Table 3 of the TOAR-II statistical guidelines provides calibrated language for describing trends and uncertainty, similar to the approach of IPCC, which allows trends to be discussed without having to use the problematic expression, "statistically significant".
Response: Thanks for taking the time to review our paper and for your helpful suggestions and comments.

Specific Comments:
1. A very important topic regarding detection of ozone trends in the troposphere is sampling frequency. Papers going back to the late 1980s have shown that low sampling frequencies (e.g. once per week ozone profiles) often fail to provide accurate monthly mean ozone values or reliable trends (Prinn, 1988; Logan, 1999; Cooper et al., 2010; Saunois et al., 2012; Chang et al., 2020; Chang et al., 2024). The modelling community is aware of this challenge (Lin et al., 2015; Barnes et al., 2016; Fiore et al., 2022) and they need long-term ozone observations with high sampling frequencies (greater than 3 times per week, if possible). The TOST product can help as it basically merges ozone observations on the regional scale, according to transport pathways, rather than through simple averaging across a pre-defined region. It would be very helpful to the modelling community if you could create a map that indicates the regions with the highest

sampling frequencies, for example, areas with three or more observations per week, and regions with 5 or more observations per week.

The panels in Figure 9 are similar to my suggestion, but I'm not sure how to interpret these plots. For example, Figure 9e shows a dark green square over Hilo, Hawaii, which seems to indicate more than 180 samples for the month of January during 2000-2009. If I divide 180 by 10 years, then I get 18 ozone samples in a month, or a sampling frequency of more than 4 times per week. Sondes are only launched from Hilo once per week, so the other samples must be due to observations associated with the forward and backward trajectories. Given that Hilo is in the middle of the Pacific Ocean, it is probably more than 4 days of transport time from the nearest ozonesonde site, and therefore any trajectory in the 5x5 grid cell above Hilo must be associated with a Hilo ozonesonde. If this is the case, then the samples in the 5x5 grid cell are not independent. The algorithm must be counting the same observation several times while the trajectory slowly traverses the 5x5 grid cell.

Is there a way for you to determine the number of independent ozone values in a 5x5 grid cell? For example, can a forward or backward trajectory from a single ozonesonde only be counted once if it falls within a particular grid cell? If you can then make a plot showing the number of independent observations within a grid cell, then it is easier to relate TOST to a sampling strategy of 1, 3 or 5 profiles per week.

Response: Thanks for the good suggestion. To determine the number of independent ozone values, we counted the forward and backward trajectories originated from an ozonesonde flying altitude only once if the trajectory falls within a particular grid cell regardless how long the trajectory stays in that cell. The number of independent samples are provided in the TOST data as well, named with a suffix of "*number_independent.asc".

The updated number of independent ozone values, for example, for January 2000s at 3-4 km and 19-20km is shown in Figure R1:

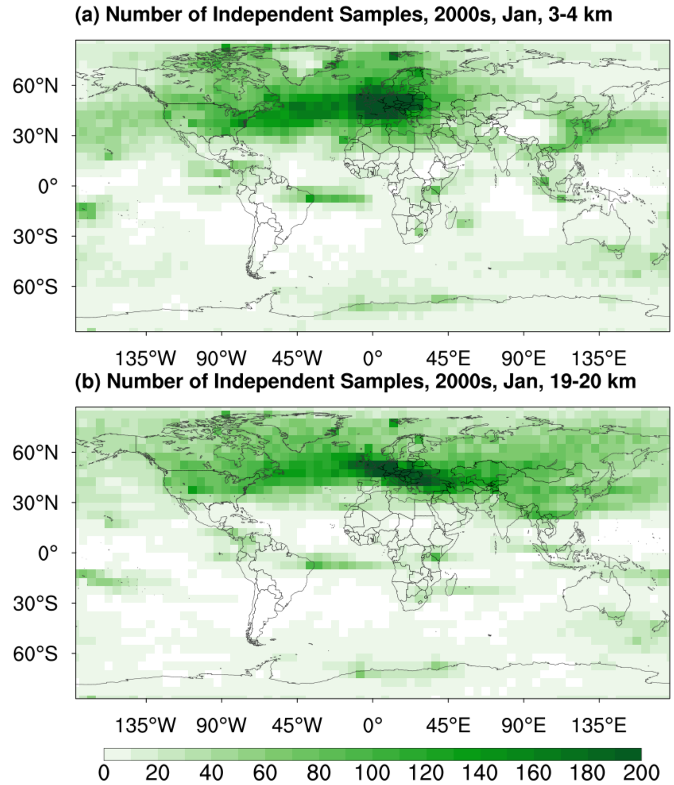

Figure R1. Global distribution of the number of independent samples for the annual mean ozone in 2000 at 3-4 km and 19-20 km.

In Hilo, the number of independent samples is 73 at 3-4 km, and 45 at 19-20 km in Jan, 2000s, which is about 4-7 samples per month (or 1-2 samples per week). The result shows that most of the samples are associated with the Hilo ozonesonde, yet still have some samples for trajectories outside the Hilo ozonesonde.

To confirm there are trajectories other than the Hilo ozonesonde to this station, we also calculated the number of independent samples for trajectories of 1-4 days. The 1-day trajectories will mostly reflect the number of samples from the ozonesonde stations, and we compare it with the 1-day trajectories generated only by the Hilo station. We found that Hilo station has 55 samples from 1-day trajectories at 3-4 km, which is the same as the 1-day trajectories generated only by Hilo station. The >1-day trajectories will mostly reflect the number of samples from other ozonesonde stations. For >1-day trajectories at 3-4 km, Hilo station has in total of 18 samples, indicating the influence of trajectories from other ozonesonde stations. Therefore, we believe the number of independent samples we calculated now is reasonable.

We also replaced the number of trajectory samples with the number of independent samples in Figure 7, so that the standard error now reflects the number of independent samples, and is correspondingly larger:

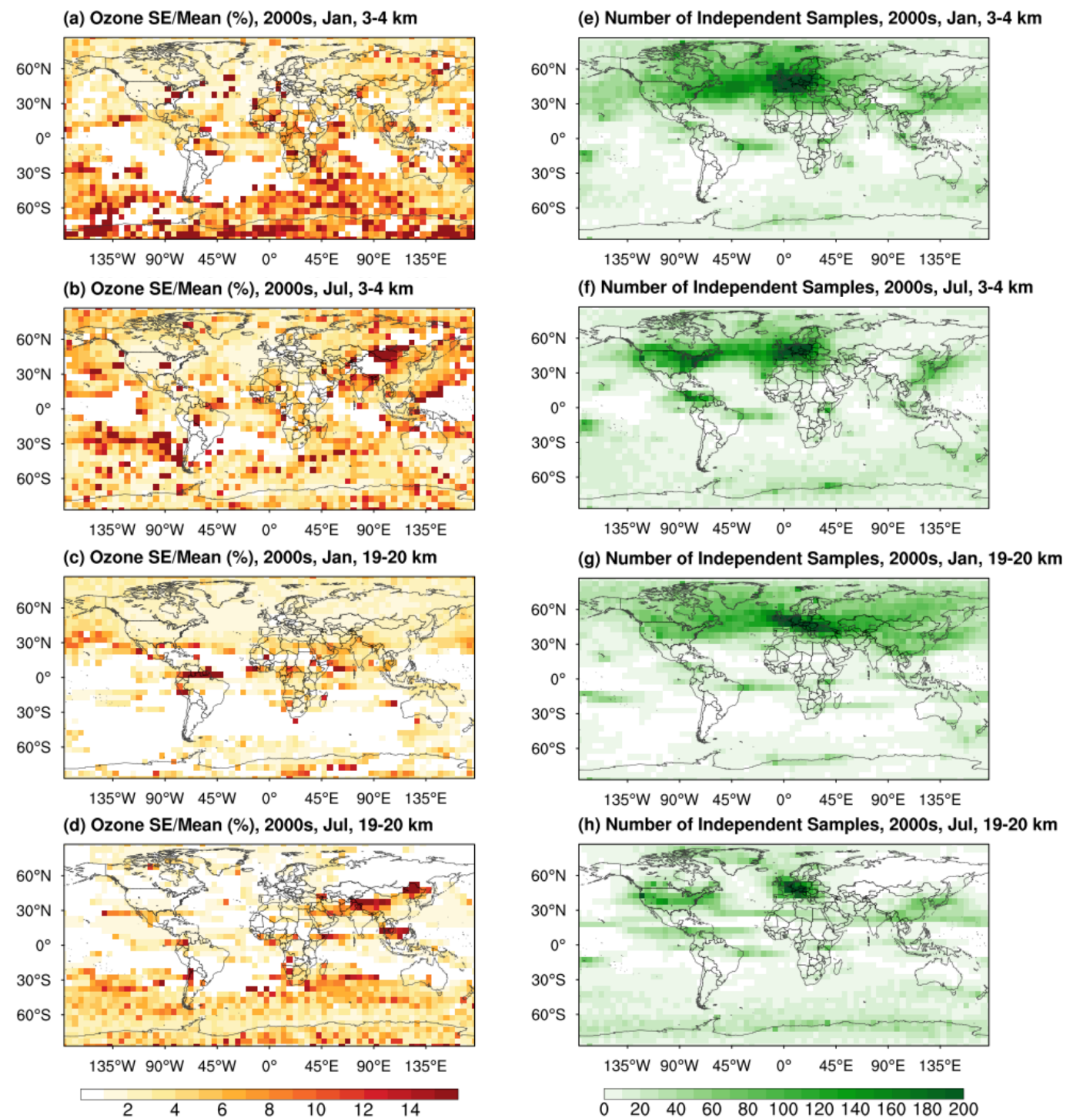

Figure 7. (a-d) Global distribution of the SE/Mean (left panels, in %) for the decadal monthly mean ozone in January and July 2000s at 3-4 km (a and b) and 19-20 km (c and d). (e-h) the same as (a-d), but for the number of independent samples in each 5 × 5° bin.

2. Lines 44-51

This introductory paragraph focuses on stratospheric ozone, while the topic of the TOAR-II Community Special Issue is tropospheric ozone. It's fine to discuss stratospheric ozone, as it impacts the troposphere, but a brief summary of the importance of tropospheric ozone is needed, especially when stating the importance of ozone for climate, as most of the radiative forcing is in

the troposphere. See Chapters 2, 6 and 7 of IPCC AR6 (Gulev et al., 2021; Szopa et al., 2021; Forster et al., 2021).
Response: Thanks for the suggestion. We have added the emphasis on the tropospheric ozone to the introduction in Line 48-52:
"Ozone is an important oxidant photochemically linked to the hydroxyl radical in the troposphere, with detrimental effects on crop productivity, natural ecosystems and human health (Fleming et al., 2018; Mills et al., 2018; Harmens et al., 2018; Vicedo-Cabrera et al., 2019). Tropospheric ozone is the third largest greenhouse gas contributing to radiative forcing, particularly in the upper troposphere (Gulev et al., 2021; Szopa et al., 2021; Forster et al., 2021)."

3. When presenting the findings from the updated TOST product the focus is on the stratosphere and there is no analysis regarding tropospheric trends. The TOST product was used by the first phase of TOAR to show that ozone has increased in both hemispheres from 1998 to 2012 (Gaudel et al. 2018). TOST was also used by IPCC AR6, and the 1998-2012 positive ozone trends are consistent with the IAGOS trends over a slightly longer period (1994-2016), as shown in Figure 2.8 below (Szopa et al., 2021). It would be helpful to provide updated tropospheric ozone trends based on TOST-v2. It would also be helpful to show the extent of the negative ozone anomalies in 2020 caused by the COVID-19 economic downturn, as previously reported by Steinbrecht et al. 2021 and Putero et al., 2023 (published in the TOAR-II Community Special Issue).
Response: Thanks for the good advice. Investigating the tropospheric ozone trend based on TOST-v2 is underway under project of "The Harmonization and Evaluation of Ground Based Instruments for Free Tropospheric Ozone Measurements (HEGIFTOM)" with careful comparison and estimates.

4. Line 26
When saying the dataset has been updated to the most recent decade (1970s-2010s) it gives the impression that the final year in the dataset is 2019, but the final year is actually 2021. Please just list the full range of the dataset using the first and final years.
Response: Revised. Thank you.

5. Lines 52-60
When reviewing the availability of long-term ozone profile records, please also mention lidar records. The lidar record at Observatoire de Haute Provence in southeastern France began in 1991; while the annual ozone anomalies from the lidar and the co-located ozonesondes differ due to sampling differences, both show a similar long-term ozone increase in the free troposphere, in the range of 1-3 ppbv/decade for 1991-2020 (Ancellet et al., 2022). Similarly, the JPL Table Mountain lidar north of Los Angeles shows an increase of 1 ppbv/decade for 2000-2023, as shown in the updated figure below (produced by Kai-Lan Chang using the method described by Chang et al., 2023). Since 2018 the Table Mountain lidar has a very high sampling frequency of 4-5 times per week. It also shows the decrease in ozone levels in 2020, associated with the COVID-19 pandemic.
Response: Thanks for the suggestion. We have added the lidar records for ozone profiles in Line 66-69:

"In addition, lidar records also provide long-term tropospheric ozone profiles, such as the Observatoire de Haute Provence lidar and the Jet Propulsion Laboratory Table Mountain lidar (Ancellet and Beekmann, 1997; McDermid et al., 2002). However, the horizontal and temporal coverages of both ozonesondes and lidars are limited by the sparse distribution of the stations (less than 100 worldwide for ozonesondes; 9 lidars from the Tropospheric Ozone Lidar Network) and their low observation frequency (1-3 times/week for ozonesondes; 1-5 times/week for lidars) (McDermid et al., 2002; Liu et al., 2013a; Chouza et al., 2019; Ancellet et al., 2022)…"

6. Line 61
In addition to providing ozone retrievals for the stratosphere and troposphere, satellites also provide total column ozone retrievals.
Response: Thanks for your point. We have revised the manuscript in Line 76-77:
"*Satellite data usually provide total column ozone retrievals that are not vertically resolved*."

7. Line 62
Satellite products can provide ozone retrievals for the lower, mid- and upper troposphere, with varying degrees of sensitivity, not just for the 6-10 km range (see section 3.3 of Gaudel et al., 2018)
Response: Thanks for the correction. We rephrased the disadvantages of satellite in observing tropospheric ozone profiles in Line 75-87:
"However, it is still challenging to retrieve tropospheric ozone through the large stratospheric ozone burden (Bhartia, 2002). Satellite data usually provide total column ozone retrievals that are not vertically resolved. The satellite ozone profiles have limited sensitivity to fine-scale atmospheric structures and the sensitivity decreases strongly toward the surface (Liu et al., 2010; Keppens et al., 2015). The direct retrieval from nadir-viewing instruments typically provides 1 independent point in the troposphere (Tarasick et al., 2019b). Large retrieval errors occur when retrieval sensitivity is low, as the solution relies heavily on the a priori (Keppens et al., 2015). In addition, single space instruments are of limited lifetime, while long-term studies on ozone require combining measurements from different instruments, which could introduce uncertainty due to the differences and drifts among datasets (Rahpoe et al., 2015)…"

8. Lines 68-72
A new area of global modelling involves the production of chemical reanalyses, which assimilate satellite data, to improve the quantification of tropospheric ozone, e.g. Miyazaki et al., 2020a,b; Colombi et al., 2021.
Response: Thanks for the suggestion. We have added the assimilation of satellite data using chemical models in Line 97-101:
"Some advanced models can improve global tropospheric ozone in 3 dimensions by assimilating satellite data to enhance the modeling accuracy (Miyazaki et al., 2020a; Colombi et al., 2021). However, in addition to the aforementioned sources of uncertainties, such assimilations still rely on the sufficiency and spatial-temporal continuity of the satellite data (Huijnen et al., 2020; Miyazaki et al., 2020b)."

9. Line 130
The Data and Methods section needs to state how the tropopause is defined, as the product is provided in terms of both the troposphere and stratosphere. If a forward or backward trajectory

begins in the troposphere and the final location of the trajectory particle, after 4 days, is above the tropopause, is this ozone observation categorized as being in the troposphere, or stratosphere?
Response: Thanks for the questions. Ozonesonde profiles are labelled according to their origin in the troposphere or stratosphere (defined by the WMO definition from the measured ozonesonde temperature profile). This allows us to generate forward and backward trajectories by only tropospheric or only stratospheric ozonesonde data to provide "troposphere-only" and "stratosphere-only" TOST fields, to be used with models, which generally calculate a tropopause. For the combined "trop_strat" product, no tropopause information is needed.
We described how we calculated the "troposphere-only" and "stratosphere-only" data and the definition of tropopause in Method in Line 232-238:
"In addition, the ozonesonde data at the 26 levels are labelled according to their origin in the troposphere or stratosphere to generate forward and backward trajectories by using only tropospheric ozonesonde data or only stratospheric ozonesonde data. The tropopause is determined following the WMO definition from the measured ozonesonde temperature profile."

10. Line 249-251
How were the IAGOS data averaged temporally? Into monthly means? What is the horizontal resolution? 5x5 degrees? How many airports were used? Did you use just the vertical profiles, or also the cruise level data? Do you have a data availability threshold? For example, do you require at least 4 profiles in a month to produce a monthly mean?
Response: Thanks for the questions. IAGOS data are averaged into monthly means with 1 km vertical resolution from sea level into horizontal bins of 5*5 degrees. In total, we used IAGOS data from 310 airports along the entire flight routines. We have not set data availability threshold when comparing the IAGOS data with TOST, i.e., the comparison is made for the grids having one or more data per month for both IAGOS and TOST.
The average of IAGOs data is rephrased with more details in Line 325-329:
"Here, the IAGOS ozone profiles were processed into 1 km layers from sea level and averaged into bins of 5° latitude and 5° longitude for each month. In total, all IAGOS ozone data from 310 airports were used for the comparison (Table S2). Then, the processed IAGOS ozone profiles were matched with the TOST ozone for the corresponding grids to examine the performance of TOST in the troposphere."

11. Line 540
The Guidance note on best statistical practices for TOAR analyses (described above) asks for all trends to be reported with 95% confidence intervals and p-values, and in units of ppbv decade-1. In the submitted manuscript trends are only reported for the stratosphere and in units of ppbv year-1. If ppbv year-1 is the standard unit for reporting ozone trends in the stratosphere, then please retain this unit, otherwise please follow the TOAR guideline.
Response: Thanks for the guidance. We have revised the trend by reporting the decadal trend with 95% confidence intervals in Line 633-634:
"There are non-significant trends in the ozone concentrations at 21-22 km (by 0.5±0.6 %/decade) and 24-25 km (by -0.2±0.9 %/decade) from 1998 to 2021"

12. Figure S3
Why compare IAGOS and TOST over the range 50-150 ppb which includes tropospheric and stratospheric samples? If a monthly mean value for IAGOS observations is 100 ppb then it is

very likely composed of both tropospheric samples (less than 100 ppb) and stratospheric samples (greater than 100 ppb). According to Figure S3d an IAGOS monthly mean of 100 ppb can correspond to a TOST value anywhere from 50 ppbv (mostly tropospheric samples) to 150 ppbv (mostly stratospheric samples). Clearly these two data sets are not sampling the same air masses and this is not an apples-to-apples comparison, so I don't see the value in these correlation plots. Compared only below tropopause.

Response: Thanks for the question. The separation of <50 ppb and 50-150 ppb here is to see the comparisons of IAGOS and TOST data in the lower and upper troposphere. However, we agree that comparing the range of 50-150 ppb could result in comparing different air masses sampling. Therefore, we only keep Figure 7 (the comparison of <50 ppb ozone samples) to make sure both IAGOS and TOST record tropospheric ozone, which is also the purpose of this figure: to compare the tropospheric ozone in TOST to another broadly trusted tropospheric ozone data (IAGOS). In addition, only the comparison of 1994-2021 was kept in Figure 7 for conciseness.

13. Section 4.2

Every year, stratospheric ozone trends and variability are updated in the Global Climate chapter of the State of the Climate reports. The most recent edition (Dunn et al., 2023) provides an update through the end of 2022. In particular, Figure 2.64 compares several products and shows stratospheric ozone levels at 22 km for the latitude band 35N-60N, similar to your Figure 13. The SWOOSH product (Davis et al., 2016) is a combined satellite product, bias corrected against ozonesonde observations and provides global coverage. How does TOST compare to these other products, and does TOST provide any new information?

Response: Thanks for the questions and comments. Although stratospheric ozone trends and variability are updated every year, the ozone trends in the lower stratosphere still have large uncertainties and differences (Ball et al., 2020; Li et al., 2023). Therefore, focusing on the lower stratosphere, we used TOST data to compare the lower stratospheric ozone trend (at 21-22km and 24-25km) since 1998 with recent studies. Over Northern Hemisphere mid-latitudes, the time series of the updated TOST shows an overall insignificant change in lower stratospheric ozone after 1998 (Figure 12), which is different from the decreasing trend reported using satellite-based data (Ball et al., 2018, 2019; Szeląg et al., 2020; Li et al., 2023). Therefore, more in-depth studies of stratospheric ozone trends, especially in the lower stratosphere, are necessary. Accordingly, we renamed Section 4.2 as "Long-term trend in the lower stratospheric ozone" and added the result in this section in Line 609-614:

"Following the implementation of the Montreal Protocol and its amendments, recent studies have found an increase in upper stratospheric ozone since the late 1990s (Chipperfield et al., 2017; Szelag et al., 2020; Dunn et al., 2023). However, the lower stratospheric ozone trend remains highly uncertain (Ball et al., 2020). Quantifying lower stratospheric ozone trends depends largely on the quality of the observational datasets (Li et al., 2023)."

And in Line 636-645:

"There is no significant trend in the ozone concentrations at 21-22 km (0.5±0.6 %/decade) and 24-25 km (by -0.2±0.9 %/decade) from 1998 to 2021, indicating little change of lower stratospheric ozone, despite the fact that 25 years have passed since peak stratospheric chlorine. Recent studies using merged satellite data suggest that the decrease in the lower stratospheric ozone is offsetting the increase in the upper stratosphere (Ball et al., 2018, 2019; Szelag et al., 2020; Li et al., 2023. However, in the Northern Hemisphere mid-latitudes, TOST indicates no significant trend in the lower stratospheric ozone after the late 1990s. The differences between

satellite-based data and TOST call for further in-depth studies on the stratospheric ozone trend, especially in the lower stratosphere."

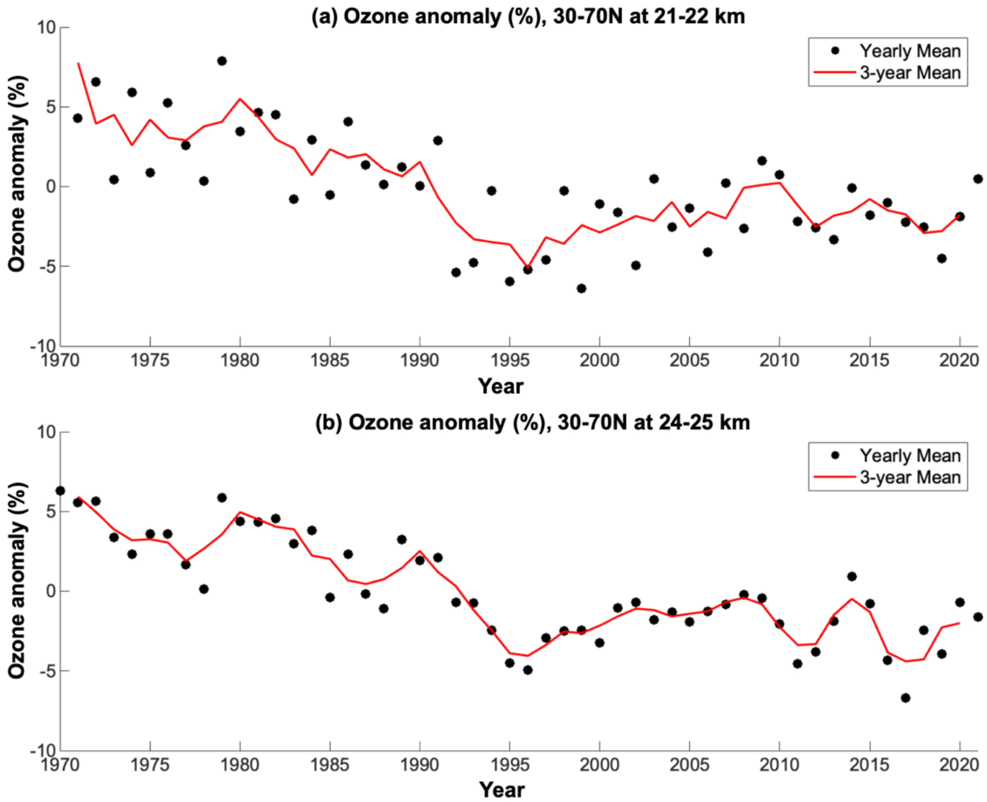

Figure 12. TOST time series of the annual mean ozone mixing ratios anomaly (in %) averaged over 30°-70°N over 21-22 km altitude (a) and 24-25 km altitude (b). The black dots represent the annual mean ozone concentrations from the area-weighted average of the grid cells over 30-70°N with ozone data throughout 1970-2021. The red line is the 3-year running mean.

References

Ball, W. T., Chiodo, G., Abalos, M., Alsing, J., and Stenke, A.: Inconsistencies between chemistry–climate models and observed lower stratospheric ozone trends since 1998, Atmospheric chemistry and physics, 20, 9737-9752, 2020.

Ball, W. T., Alsing, J., Staehelin, J., Davis, S. M., Froidevaux, L., and Peter, T.: Stratospheric ozone trends for 1985–2018: sensitivity to recent large variability, Atmospheric Chemistry and Physics, 19, 12731-12748, 2019.

Ball, W. T., Alsing, J., Mortlock, D. J., Staehelin, J., Haigh, J. D., Peter, T., Tummon, F., Stübi, R., Stenke, A., and Anderson, J.: Evidence for a continuous decline in lower stratospheric ozone offsetting ozone layer recovery, Atmospheric Chemistry and Physics, 18, 1379-1394, 2018.

Harmens, H., Hayes, F., Mills, G., Sharps, K., Osborne, S., and Pleijel, H.: Wheat yield responses to stomatal uptake of ozone: Peak vs rising background ozone conditions, Atmospheric environment, 173, 1-5, 2018.

Li, Y., Dhomse, S. S., Chipperfield, M. P., Feng, W., Bian, J., Xia, Y., and Guo, D.: Quantifying stratospheric ozone trends over 1984–2020: a comparison of ordinary and regularized multivariate regression models, Atmospheric Chemistry and Physics, 23, 13029-13047, 2023.

Szeląg, M. E., Sofieva, V. F., Degenstein, D., Roth, C., Davis, S., and Froidevaux, L.: Seasonal stratospheric ozone trends over 2000–2018 derived from several merged data sets, Atmospheric Chemistry and Physics, 20, 7035-7047, 2020.
Vicedo-Cabrera, A. M., Sera, F., Liu, C., Armstrong, B., Milojevic, A., Guo, Y., Tong, S., Lavigne, E., Kyselý, J., and Urban, A.: Short term association between ozone and mortality: global two stage time series study in 406 locations in 20 countries, bmj, 368, 2020.