Review 2 by Michael Prather

You already have some excellent suggestions from RC1. My issues mainly focus on the methodology and future of these ozone data sets.
This paper documents and presents a long-term global, gridded harmonized ozone data set for the troposphere and lower stratosphere (TOST) that is readily amenable for developing model metrics and studies of trends and interannual variability. It is very well written for the most part and will be a valuable addition to a broad community studying atmospheric chemistry and transport, global air pollution, and global change. The core datasets are the ozonesondes, and MOZAIC/IAGOS is used for validation – a great choice for well calibrated, highest resolution possible atmospheric composition measurements. The updated TOST-v2 is a great product. Yet, this is a disappointing paper in merely repeating the TOST-1 protocol without much thought as to the use of the data in modern models. At this point it needs to be published as is (with some minor noted corrections) but with the added recognition/recommendations of how to do it better.
Response: Thanks for taking the time to review our paper and for your helpful suggestions and comments.

1. To me, the obvious question here is: why not include the IAGOS data as a source for TOST? It seems like you are wasting a major resource by using it only for validation. I am not asking you to create TOST-3 for this paper, but at least you could discuss this at the beginning. Are there fundamental problems with this? or just too much work for now (that is OK).
Response: Thanks for this question. In fact, we started to do this quite some time ago and quickly became aware that there was a persistent bias between IAGOS and sonde measurements.
The result of that is a recent study from Wang et al. (2024), which evaluated the agreement between IAGOS and ozonesonde data, resulting in all sonde types showing significant average biases with respect to IAGOS (higher by 5-10% than IAGOS), and the relative bias increasing modestly with altitude. This result also agrees well with our Figure 6, comparing IAGOS and TOST data (the relative difference is 5.84%). In addition, the time periods of IAGOS and ozonesonde data are not generally the same, so merging these two data types can introduce spurious jumps in timeseries, if the bias is not properly resolved.
At present, a similar trajectory-derived ozone climatology based on IAGOS data is now under construction in our group and will be published as an independent dataset like TOST.

2. Abst. "of 5° × 5° × 1 km (latitude, longitude, and altitude" sound nice but it is missing two important quantities: (1) is "altitude" really just altitude (km above the surface) or is it "pressure altitude" ? be specific (log p, or US STD atmos p like flight levels); (2) time is critical here, what is the resolution and method of averaging? I see in L164 that you used monthly averages, please state this up front.
Response: Thanks for the question, the altitude in the previous version is the geometric altitude above the surface, and now we have added the pressure altitude as suggested. The time is at two resolutions: annual mean and decadal-monthly mean.

Oh, now I see in L204 ("The resulting ozone fields are given in two altitude coordinates (altitude above sea level and altitude above ground level) for users' convenience") that you are using geometric altitude. This is really problematic since the altitude of the land surface depend

heavily on the resolution of the model you (and your users) are using. I think these are possibly the worst possible vertical coordinates you could use, especially for the 6-26 km region where the results are most reliable. The use of altitude requires one to know the temperature profile, which is seriously problematic since any model profile may NOT be what you use and there fore cannot be compared. If you are using a fixed T profile, then just provide the data set in pressure coordinates.
I think the data set must really be in pressure coordinates to be useful to any 3D model. This you can and should fix.
Response: Thanks for the good suggestion. We have added the TOST-v2 in pressure altitudes, including the altitude above sea level and altitude above ground level. We also mentioned this improvement in the Method (Line 369-371), Result (Line 1071-1072) and Conclusion (Line 1316-1388). In this study, we used geometric coordinates as the example for all the comparisons, validations and investigations.
We also described how we produced TOST in pressure altitudes in Method in Line 225-230:
"To produce the mapping with pressure altitudes, we also averaged the 4-day backward and forward trajectories in bins of 5° latitude and 5° longitude for every month, using the pressure altitudes generated by HYSPLIT trajectories. The 26 pressure altitudes are 950, 850, 750, 650, 550, 450, 400, 350, 300, 250, 225, 200, 175, 150, 125, 100, 90, 80, 70, 60, 50, 40, 35, 30, 25, 20 hPa, which is determined and adjusted based on the ERA5 pressure coordinates and the 1-26km geometric altitudes."

3. Overall big problem and opportunity – may be insurmountable, but should be recognized. Spatio-temporal averaging destroys the ozone structure anywhere near the tropopause. It is clear that this data set does not resolve tropopause ridges-troughs nor strat-trop folds – therefore the averaging of mole fraction ozone means that stratospheric ozone dominates the abundance well into the troposphere. You simply average the ozone mole fraction in your large cells over the month. It would be great to produce a more nuanced data set that considers the natural variability in ozone. Specifically, why not give 10-25-50-75-90 %iles, that way one can test the high resolution (no serious models are running % deg resolution anymore), high-frequency simulations. These statistics would help identify the frequency of strat-vs-trop, etc. and make model comparisons with the coarse resolution you use more informative. I think you should be more expansive in diagnosis.
Response: Thanks! Please note that we provide three versions of the TOST fields, based on the origin of the ozonesonde data (tropospheric or stratospheric, defined by the WMO definition from the measured ozonesonde temperature profile): "troposphere-only" and "stratosphere-only" and a combined "trop_strat" product. The last is indeed subject to the issues you note.
To study the variations of ozone, we have provided the standard deviation of ozone trajectories of each grid. As suggested, we now also added the annual and decadal-monthly 25-50-75 %-iles ozone mapping in TOST-v2. Because the number of trajectories could be limited, 10 and 90%-iles are not provided.
We also noted the meaning of providing the percentiles of ozone in Method at Line 244-246:
"In TOST-v2, we also generate the corresponding datasets that show ozone variation at 3 percentile levels (25, 50 and 75%)."
In Result at Line 554-555:
"Furthermore, TOST-v2 provides additional information that shows ozone variations in 3 percentile levels (25, 50 and 75%)."

And in Conclusion at Line 671-672:
"In addition to the seasonal, annual or decadal-monthly means, the corresponding datasets for ozone variations at 3 percentile levels (25, 50 and 75%) are also provided."

4. L61: The satellite data indeed have trouble with the troposphere (except with product involving cloud slicing or OMI-MLS as in Ziemke et al). I am even worried that MLS and SAGE may have difficulties in the UT/LS give the resolution you cite.
Response: Thanks for the question. Both SAGE and MLS are designed for measuring stratospheric ozone. It is recommended to use MLS ozone profiles only above 261 hPa (Livesey et al., 2022). We compared the MLS and SAGE profiles in Figure 6 using only >16km, which is even higher than the recommended altitude (~10 km) and should avoid comparing the too-large bias in the UT/LS area.

5. L77-79: The argument for ozone being inert for 4 days along the trajectory is reasonable for the UT/LS, but the out-of-date Jacob (1999) paper you use here is simply wrong for the lower troposphere. Look at the regions of intense ozone loss (>5 ppb/day) in the ATom transects (Prather, Guo, Zhu 2023, doi: 10.5194/essd-15-3299-2023) or the 3-5 day perturbation lifetime of surface ozone pollution in Prather & Zhu (2024, Lifetimes and timescales of tropospheric ozone, Elementa, doi: 10.1525/elementa.2023.00112). I do not think you can easily do anything (or even should do anything) about this for your TOST-2 product, but there should be a recognition of the potential error.
Response: Thanks for pointing out this out. Our results did show the bias from assuming a 4-day lifespan of ozone in the lower troposphere. For example, in Figure 3, the surface (boundary layer) ozone shows a positive bias of the median, in all decades, of up to 12%. In addition, in Figure 4, the larger discrepancies are shown near the planetary boundary layer due to the fact that a 4-day lifespan for ozone could be unreal for the lower troposphere. In the uncertain analysis, we emphasized that surface ozone could be more biased than other altitudes.
In this version, we have cited the study of Han et al. (2019) and Prather & Zhu (2024) in the introduction. In Han et al. (2019), the lifetime of ozone at the middle troposphere (500 hPa) and the surface is estimated to be >10 days and 1.1-11.3 days, respectively. Therefore, the extension of the 4-day lifespan for ozone is generally reasonable for generating the TOST data.
In future studies, we will improve the TOST in near-surface by using varied trajectory length for different altitudes of the atmosphere according to their mean lifespan.

6. L169: The new HYSPLIT may be numerically accurate but the NCAR/NCEP wind fields seem totally out of date – the vertical resolution (17 layers from 0 to 32 km = 2 km at best near the tropopause) can hardly resolve vertical motions in the UT/LS. Why not use more modern fields like ERA-5 or MERRA-2? It makes the paper look lazy, you updated the sondes, but just ran with the old parts of TOST-1. I know you cannot fix this, but it should be recognized as a problem (like the minimal use of IAGOS observations) that should be upgraded in TOST-3.
Response: Thanks for the suggestion. We agree that updated wind fields could improve TOST accuracy. While the 17 layer NCEP fields do lack vertical resolution, they were until fairly recently the only reanalysis dataset that offered consistency back to the 1960s, when our sonde data begin. Other NCEP data are for more recent years. MERRA starts in 1980. The ERA5 dataset would be an obvious improvement, but the effort involved in switching is considerable.

7. L272: I was going to congratulate the authors on their correct use of nmol/mol as the measure of ozone abundance and then I hit the incorrect use of 'ppbv' ("RMS of 21.1 ppbv, and higher bias (2.9 ppbv) and"). The 'by volume' should have been scoured out of this community by now but many prominent colleagues continue to abuse this. The 'volume' is not mole fraction since virial corrections would need to be applied, and most all measurements calibrate to dry air mole fraction.
Response: We meant ppb. Thanks for catching this.

8. L475: You really should be comparing tropospheric O3 column (DU or mean ppb) with Ziemke et al's work. The whole paper is well referenced within its limitations (noted above), but you simply must compare the features in Figures 8 and later with Ziemke's work.
Response: Thanks for the suggestion. This is not as simple as you might think, since many TOST columns are missing data at one or more levels. We are working on a better gap-filling estimation technique, and in an upcoming paper, we have compared the tropospheric O3 column with Ziemke et al's OMI data. Hope to submit it soon.

9. L555: Again, note that this is monthly averaged.
Response: Thanks for pointing this out. The data is provided at three temporal resolutions: seasonal, annual and decadal-monthly mean, and we have emphasized this in the conclusion. The reason we could not provide monthly-mean data is that despite the trajectory filling, monthly-mean data still have large gaps. Therefore, to increase the data availability, we provide the seasonal, annual and decadal-monthly mean data, which can be used for spatial analysis.

References

Han, H., Liu, J., Yuan, H., Wang, T., Zhuang, B., and Zhang, X.: Foreign influences on tropospheric ozone over East Asia through global atmospheric transport, Atmospheric Chemistry and Physics, 19, 12495-12514, 2019.

Livesey, N., Read, W., Wagner, P., Froidevaux, L., Santee, M., Schwartz, M., Lambert, A., Millán Valle, L., Pumphrey, H., and Manney, G.: Earth Observing System (EOS) Aura Microwave Limb Sounder (MLS) version 5.0 x level 2 and 3 data quality and description document Version 5.0–1.1 a (Tech. Rep.), Jet Propulsion Laboratory, California Institute of Technology. Retrieved from https://mls. Jpl. Nasa. Gov/data/v5-0_data_quality_document. Pdf, 2022.

Prather, M. J. and Zhu, X.: Lifetimes and timescales of tropospheric ozone: Global metrics for climate change, human health, and crop/ecosystem research, Elementa: Science of the Anthropocene, 12, 2024.