Review 1

The paper describes an updated and improved version of the Trajectory Mapped Ozonesonde Dataset (TOST), which provides gridded ozone profile data from the 1970s until 2020. Overall this appears to be a good dataset. The paper is overall OK and suitable for ACP (or even better ESSD?). However, there are a number of issues that should be addressed before publication.

The paper is very long and contains a lot of redundant information in text and plots. Further down I suggest a number of way to simplify Figures. I strongly suggest to also shorten the corresponding text and to shorten and focus the conclusions section.
Response: Thanks for taking time to review our paper and for your helpful suggestions and comments.

1. It appears the TOST data set uses a 5°x 5°x 1km latitude x longitude x altitude grid (e.g. lines 26, 183, 184). However, it is not really clear what the provided time coordinate is. From lines 189 and 190 it appears that one time coordinate might be 12 monthly means, for each of the 5 decades 1970 to 1979, 1980 to 1989, ..., 2010 to 2019. Another time coordinate seems to be 52 annual means for each of the years 1970 to 2021. This should be clarified in a few places, especially in Abstract and Conclusions. Also, it begs the question, why the data-set is not simply provided as 12 monthly means for each of the 52 years.
Response: Thanks for the points. The data are provided at three temporal resolutions: seasonal, annual mean and decadal-monthly mean. This information is now explicitly provided in Abstract, Conclusions, and text
In Line 24-27 in Abstract:
"Here, the seasonal, annual and decadal-monthly Trajectory-mapped Ozonesonde dataset for the Stratosphere and Troposphere (TOST) ozone climatology is improved and updated from 1970-2021 on a grid of 5° × 5° × 1 km (latitude, longitude, and altitude) from the surface to 26 km by geometric and pressure coordinates"
In Line 660-663 in Conclusions:
"Similar to TOST-v1, the ozone in each season, in each year (1970-2021) and in each month of a decade (January to December from the 1970s to 2010s) are provided in 3-dimensional grids of 5° × 5° × 1 km (latitude, longitude, and altitude)."
We also summarize the coordinate systems, starting levels, temporal resolution, and data types for TOST data in Method (Line 239-244) and Table 1:
"Based on this mapping, TOST-v2 was generated at 26 altitude levels in two altitude coordinates (by geometric levels and pressure levels), from two altitude starting levels (altitude above sea level and altitude above ground level), for three temporal resolutions (in the seasonal mean for each year, the annual means for each year from 1970 to 2021, and monthly means for each decade from the 1970s to the 2010s) and with three types of data fields (trop-strat, troposphere-only and stratosphere-only) for users' convenience (Table 1)."

Table 1. The classifications and descriptions for TOST data on coordinate systems, starting levels, temporal resolution, and data types.

Main Classifier	Type	Description
1. Coordinates	Geometric	Altitude coordinates are 1, 2, …, 25, and 26 km at 1-

		km vertical resolution.
	Pressure	Altitude coordinates 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.
2. Starting levels	Sea-level	Data start at the altitude of the sea surface. Values for levels below the ground surface are set to null.
	Ground-level	Data start at the altitude of the ground surface.
3. Temporal resolutions	Seasonal	Data are the mean for each season of the year (1970–2021).
	Annual	Data are the mean for each year (from 1970-2021). Each grid requires at least one value per season to be included in the annual data.
	Decadal-monthly	Data are the mean for each month of a decade (from the 1970s to 2010s).
4. Ozonesonde data used	Trop_Strat	Data are based on ozonesonde profiles from both the troposphere and stratosphere.
	Troposphere-only	Data are based on ozonesonde profiles only from the troposphere.
	Stratosphere-only	Data are based on ozonesonde profiles only from the stratosphere.
5. Ozone variations	Mean	The mean trajectories for each grid point over a period (a month, a year, or a month of a decade).
	25th, 50th and 75th percentiles	The 25th, 50th and 75th percentiles of the trajectories for each grid point over a period (a month, a year, or a month of a decade).
6. Supplement data	Smoothed data	Smoothed ozone fields by fitting the maps at each level with a linear combination of spherical functions (only for decadal-monthly and annual data).
	N	The number of trajectories passing through the grid cell.
	Number of independent samples	The trajectory was counted only once in a grid cell when the trajectory passes that cell regardless of how long the trajectory stays in that cell.
	Var	The variability of the trajectories values in each grid cell.
	SE	The ratio of the Var to the square root of the number of trajectories in each grid cell.
	SD	The ratio of the Var to the mean of trajectories values in each grid cell.
	SE/Mean	The ratio of the SE to the mean of trajectories values in each grid cell.

The reason we could not provide monthly-mean data is because, 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.

It is worth noting that all validation is carried out at monthly time step at individual stations or by regional mean.

2. To me, the paper contains way to many similar plots and panels. This makes it very hard for a reader. If there is no significant difference between seasons, decades, ... just show one plot / panel. See e.g. my comment on Fig. 5 below. Additional plots could go to the supplement, but even there: If there is no significant difference between seasons, decades, ... just show one plot / panel. The goal of the paper should be to clearly bring out the major messages, not to overwhelm and confuse the reader with redundant information.
Response: Thanks for the suggestion. We have revised Figures 5, 6 and 9 to clarify and simplify the contents. Please see more details in the following responses.

3. I am quite confused by the various relative and absolute measures used for differences in the validation part of the paper. Sometimes the authors seem to use mean relative difference (RD), sometimes bias (=absolute mean difference?) , sometimes root mean square differences (RMS, absolute or relative?), sometimes root mean square differences of the mean (RMS/sqrt(N), absolute or relative?). I think this should be clarified, and if at all possible simplified and unified.
Response: Thanks! Apart from correlation coefficient (R) and linear fitting coefficient, we used relative difference (RD) and normalized root mean square (NRMS) to represent the **relative** difference between the two compared data, and used bias and root mean square (RMS) to show the **absolute** difference between the two compared data.

The relative difference allows comparing uncertainties and accuracies for TOST ozone estimations at different altitudes where ozone concentrations vary greatly. We have added the detailed equations for these metrics in Section S1 in the supplement file:
"*Multiple metrics were used to indicate the agreement and differences between the TOST (use y here) and ozonesonde/aircraft data (use x here).*

1. Correlation Coefficient (R, unitless):

$$R = \frac{\sum(x_i-\bar{x})(y_i-\bar{y})}{\sqrt{\sum(x_i-\bar{x})^2\sum(y_i-\bar{y})^2}}$$, *where* $\bar{x}$ *and* $\bar{y}$ *is the mean of the x and y variables, respectively.*

2. Linear fitting coefficient (m, unitless), with the intercept set to 0:

$$m = \frac{\sum(x_i y_i)}{\sum({x_i}^2)}$$

3. Bias (in ppb):

$$Bias = \frac{1}{n}\sum(x_i - y_i)$$

4. Relative Difference (RD, in %):

$$RD = 100 \times \frac{Bias}{\bar{x}}$$, if comparing with ozonesonde data *x*

$$RD = 100 \times \frac{Bias}{\frac{1}{2n}\sum(x_i+y_i)}$$, if comparing with satellite data *x*

5. Root Mean Square (RMS, in ppb):

$$RMS = \sqrt{\frac{1}{n}\sum(x_i - y_i)^2}$$

6. Normalized Root Mean Square (NRMS, in %):

$NRMS = 100 \times \frac{RMS}{\bar{x}}$, if comparing with ozonesonde data x

We have stated the metrics and their definitions in Method (Line 269-276):
"Multiple metrics were used to indicate the agreement and differences between TOST and other data. We used correlation coefficient (R) to present the agreement of the two compared datasets, and linear fitting coefficient with the intercept set to 0 to show the overall tendency of overestimation/underestimation. We also used relative difference (RD) and normalized root mean square (NRMS) to represent the relative difference between the two compared data, and used bias and root mean square (RMS) to show the absolute difference between the two compared data. Details of the metrics can be found in Section S1."

3.1 One such confusing example is Table S2, where I have no clue in what units the various quantities are given. I assume RMS is in ppbv, which is kind-of meaningless because ~400 ppbv would be a huge 400% uncertainty in the troposphere, and a reasonable 10% uncertainty in the stratosphere. I also assume that bias is in ppbv (absolute difference), and is essentially the same as RD (which seems to be relative difference in %). If relative and absolute difference ar given (RD and bias?), why are not also relative and absolute RMS given? In Figure 2, there is a sensible separation between tropospheric, stratospheric and intermediate ozone regimes. Why is that not done here in Table S2?
Response: Sorry for the confusion. We have now added units in the original Table S2 (now Table 3). We have used the "RMS and NRMS" in corresponding to the "RD and bias". Because the comparisons between two satellite data and TOST with ozonesonde data are only for >16 km, there is no separation for the three ozone regimes for Table S2 (now Table S3).

4. Line 363 and following: What is RMSE? Not defined. I assume it is root mean square error. How is that different from RMS difference?
Response: Yes, it is RMS, sorry for the confusion. Only RMS and NRMS are used in this revision. Please see our response to Comment 3.

5. Line 460, 461: What is NRMSE? Needs to be defined. It seems to be the same as relative root mean square error / difference
Response: NRMS is defined as the root mean square divided by the mean value of the variable. Please see our responses to Comment 3 for details.
Why is it not in %?
Response: Yes, NRMS should be in %.
In most other places relative differences and relative uncertainties are in % (and absolute ones in ppbv). Please define better and make consistent, e.g. always give RD and RMS in % and ppbv.
Response: We have unified the metrics and given RD and NRMS in % to indicate they are relative differences, and given bias and RMS in ppb to indicate they are absolute differences.

6. Figure 2: I find the vertical bars for R quite confusing. I would much prefer a third set of symbols / lines.
Response: For Figure 2, we have plotted the metrics separately to avoid confusion in this revision. To simplify the metrics, only R and RD are shown now.

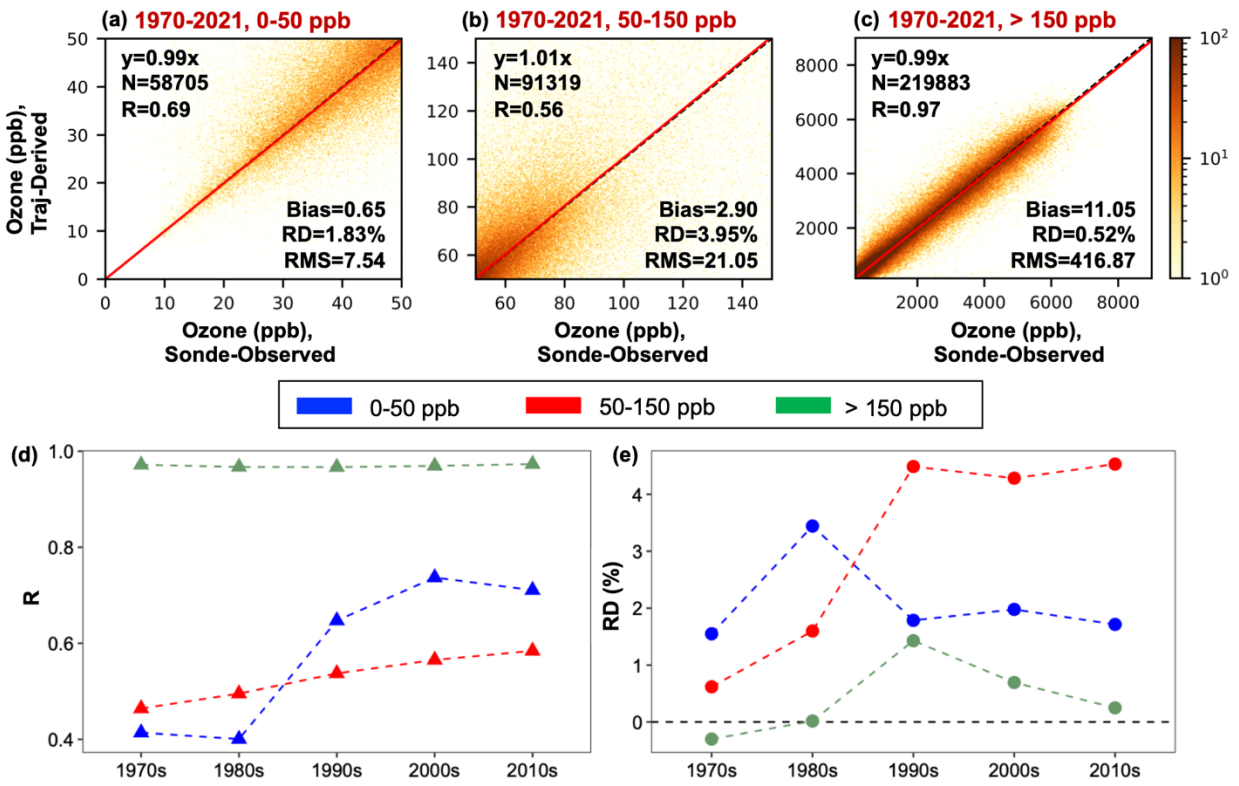

Figure 2. (a-c) Comparison of monthly average tropospheric ozone mixing ratios from ozonesondes (Sonde-Observed) and trajectory-derived TOST data (Traj-Derived) for the entire study period of ozone concentration at 0-50 ppb, 50-150 ppb and >150 ppb. Solid red lines represent the linear fitting line (with the intercept set to 0) and dashed black lines denote the 1:1 axis. N is the total number of data points, R is the correlation coefficient, Bias is the overall average difference in monthly mean values [Traj-Derived ozone - Sonde-Observed ozone, in ppb], RD is the relative difference in % [100 × (Traj-Derived ozone - Sonde-Observed ozone)/ Sonde-Observed ozone)], and RMS is the root mean square difference in ppb). Note that Traj-Derived ozone at each station is derived without input from the station itself; that is, Traj-Derived represents an ensemble of 141 separate computations of TOST, each one withholding a single validation station. (d-e) the R and RD between the Traj-Derived ozone and Sonde-Observed ozone by decade. The dashed line in (e) denotes where the RD is 0.

I assume that each dot corresponds to one latitude-longitude-altitude grid-cell and one annual mean? Should probably be stated somewhere.

Response: Each dot in Figure 2a-c represents the monthly mean ozone value in one latitude-longitude-altitude grid cell. We have mentioned in the first paragraph of 3.1 in Line 333-335: *"First, we show the overall comparison in monthly mean ozone profile between ozonesonde and trajectory-derived values without the inputs of the stations being tested (Traj-Derived), from all the existing stations at all altitude levels…"*

We also stated it again in Line 350-352:

"Each dot in Figure 2a-c represents the paired ozone concentrations from the Traj-Derived and Sonde-Observed values in each month at each latitude-longitude-altitude grid-cell, and the color indicates the density of dots…"

7. Figure 3: Why not also give numbers for the spread / width of the distributions, e.g. full-width at half maximum, or 1 standard deviation? I assume that the underlying data points are one latitude-longitude-altitude grid-cell and twelve calender months? Should probably be stated somewhere.

Response: Thanks for the good advice. We indicated the width of 1 standard deviation using thick red lines for the RD in each level and gave the value in red. Also, because the peak density values are not the focus of this plot, we deleted the values and kept only the points indicating where the density peaks.

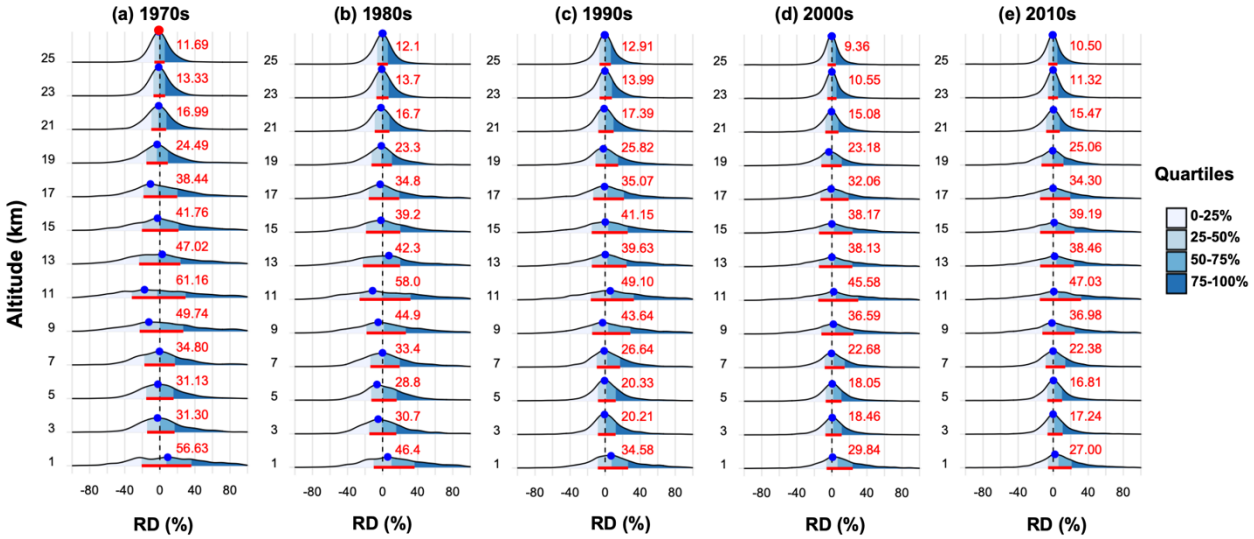

Figure 3. The relative difference (RD) of the monthly ozone mixing ratios between ozonesonde and Traj-Derived data by altitude in the 1970s, 1980s, 1990s, 2000s and 2010s, respectively. The frequency distribution of RD at every other altitudes is shown (y-axis: frequency in %, x-axis: RD in %), with the colors denoting the 4 quartiles of RD. The dashed line indicates zero difference in RD. The blue dot represents the maximum frequency. The thick red lines denotes the width of distribution at 25-75%-ile, with the corresponding width of the distribution value in red.

The data to calculate the RD distribution is from the monthly mean ozone data between ozonesonde and Traj-Derived data from all the existing stations at selected altitude levels. We have emphasized the data points for RD distribution in Line 387-389:

"The RD distributions are based on the monthly ozone concentration difference between the actual ozonesonde and Traj-Derived data from all the existing stations at the corresponding altitude level and decade."

8. Figure 5: I don't see any clear or significant differences between the top four panels, or between the bottom four panels. Therefore, I strongly suggest to just have one panel showing SAGE - TOST (all seasons, years), and one panel showing MLS - TOST (all seasons, years). It would, however, be helpful to also plot the relative RMS differences.

9. Figure 6: There is a lot of redundancy between Fig. 6 and Fig. 5. The single profile panels of Fig. 6 contain more or less the same information as Fig. 5 (especially if my suggested reduction is done). The main additional information in Fig. 6 is the seasonal variation (which is clearly

visible for MLS). Maybe there is no need for Fig. 5, or the single profile panels of Fig 6. could be dropped?
Response: Thanks. Figure 5 is now removed. Please see our responses to Comment 6.

10. Figure 7, Figure S3: Again, I don't see the need for four panels, as I don't see a significant difference between the panels. On the other hand the split between < 50 ppbv and 50 to 150 ppbv seems very artificial here. It seems to me that just one panel that includes all data from 0 to 150 ppbv would be enough and more sensible here.
Response: Thanks for the suggestion. The separations of <50 ppb and 50-150 ppb here are to see the comparisons of IAGOS and TOST data in the lower and upper troposphere. However, since ozone of 100-150 ppb can have a tropospheric or lower stratospheric source, the latter comparison is less meaningful. To make the comparison apples-to-apples, we only keep Figure 7 (the comparison of <50 ppb ozone samples) to make sure both IAGOs and TOST ozone are from tropospheric air, which is also the purpose of this figure: to compare the tropospheric ozone from TOST to another broadly trusted tropospheric ozone data (IAGOs).

11. Line 441 and following: SE/mean is that not simply the relative RMS/sqrt(N) (in %). Another example where a more consistent nomenclature and use of relative and absolute differences would be helpful.
Response: Thanks for your points. SE/Mean here is the ratio of the standard error to the mean, while standard error is standard deviation/sqrt(N). Therefore, SE/Mean is not the RMS/sqrt(N) (please see the definition of RMS in our response to Comment 3). SE indicates a confidence interval for the mean averaged over all trajectories. We use "SE/Mean" for Figure 9 to provide a relative measure of the SE, to avoid any confusion due to the magnitude differences in ozone concentrations at different altitudes.

12. Figure 9: unless there is a large and significant seasonal variation: two rows might be enough. However, I would like to see a third column with relative RMS (in %, without the 1/sqrt(N)). I guess this third column would carry comparable information as Fig. 10?
Response: Thanks for the suggestion. The seasonal variation here is to show the uncertainties of TOST in different seasons, which is clear at the 3-4km altitude level that the warm season has higher SE/Mean than the cold season. Also SE/Mean varies less with season in the stratosphere than in the troposphere. Therefore, we still keep the seasonal variations in the figure.
If without the 1/sqrt(N), we present the coefficient of variance (CV, in %) here (Figure R1), which is calculated as the ratio of standard deviation to the mean value. Compared to SE/Mean, CV has a relatively larger value than SE/Mean, as 1/sqrt(N) is not included. For the reasons described in our responses to Comment 11, we have kept SE/mean in the figure.

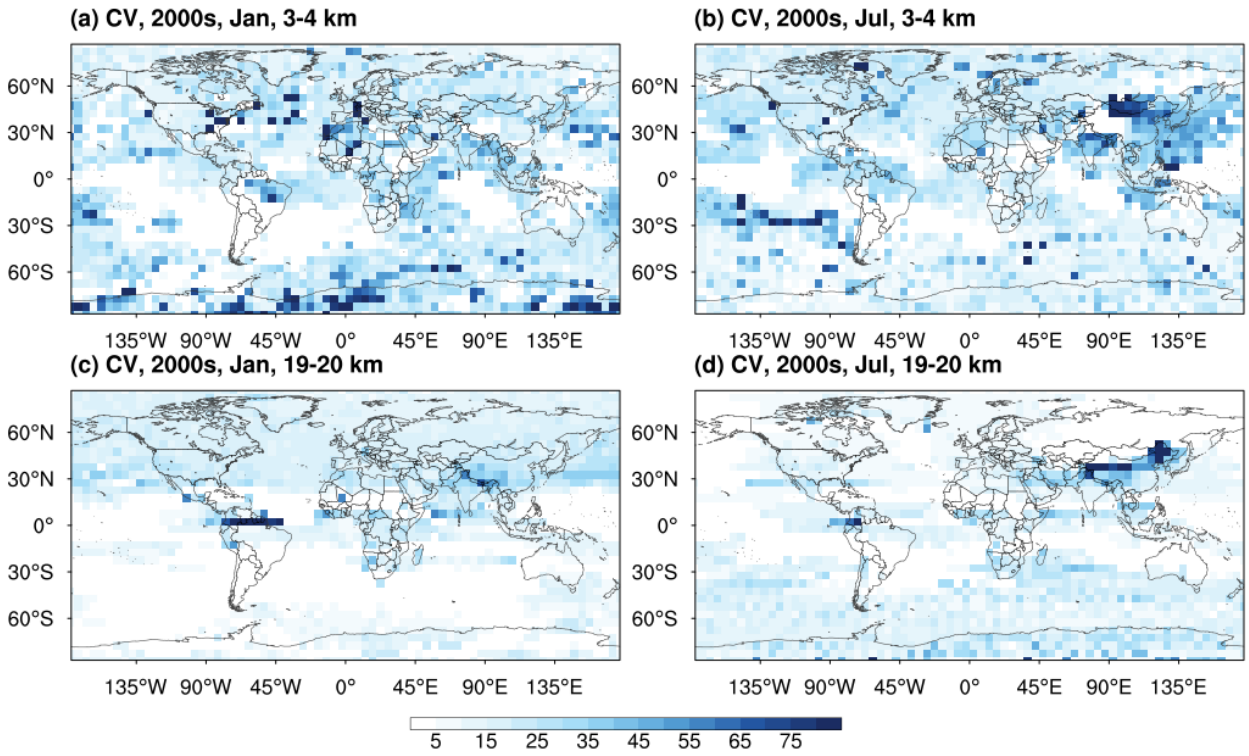

Figure R1. Global distribution of the coefficient of variance (CV, in %) for January and July 2000s at 3-4 km (a and b) and 19-20 km (c and d).

These RMS numbers should be compared with estimates of ozone sonde uncertainty, e.g. those given be Tarasick et al. 2016, 2021.
Response: Note that the calculated RMS is a bias, while the "uncertainty" used by Tarasick et al. (2016, 2021) is random uncertainty from sources in the sonde instrument and measurement. We use the differences between Traj-Derived ozone and ozonesonde data to estimate the uncertainty in the TOST product, This is substantially larger than the (order ~5%) sonde measurement uncertainties discussed in Tarasick et al. (2016, 2021).

13. Figure 10: Would not Fig. 10 and this entire uncertainty discussion (section 3.5 and Figs. 9 and 10) fit much more logically directly after Figs. 3 and 4 and section 3.1, which also compares Traj-Derived with Sonde??
Response: Each panel in Figure 10 is meant to investigate how NRMS changes with altitudes, seasons, latitude zones and decades to give readers a clear view of where and when the uncertainty of TOST could be higher. This is not only the comparison between Traj-Derived and Sonde-observed ozone, but also an important caveat for users of TOST to know where and when the data would have higher uncertainties. Therefore, we hoped that putting this uncertainty discussion after the comparisons and validations of TOST data would serve as a good summary and caveats for the first part of this paper (validations and comparisons of TOST).

Is NRMSE not the same as relative RMS? Should it not also be given in %.
Response: Yes, we have unified this metric as "NRMS" in the manuscript, and the unit of NRMS is given in %. Please see our responses in detail to Comment 3.

Should panel a.) not also have altitude on the vertical coordinate, like all the other plots?
Response: To make the panels consistent (NRMS on the vertical coordinate and altitudes/seasons/latitude zones/decades on the horizontal coordinate), it is better to keep the altitudes on the horizontal coordinate.
These RMS numbers and the profile in panel a.) should be compared with estimates of ozone sonde uncertainty profiles, e.g. those given be Tarasick et al. 2016, 2021.
Response: As noted above, the "uncertainty" used by Tarasick et al. (2016, 2021) is random uncertainty from sources in the sonde instrument and measurement. These RMS numbers are substantially larger than the (order ~5%) sonde measurement uncertainties discussed in Tarasick et al. (2016, 2021).

14. Lines 503 to 505: This is important and needs to appear prominently also in the conclusions, and in the introduction (e.g. after line 91). We don't need another "tropical ozone hole" paper and consequent rebuttal like Chipperfield et al. 2022.
Response: Thanks for the suggestion. We have emphasized this incorrect use of TOST in the introduction in Line 125-128:
"For users' convenience, the remaining gaps after trajectory mapping were further filled with a linear combination of spherical functions and provided as "smoothed" data in TOST-v1. Yet, the smoothed data should be used with caution; otherwise, misinterpretation of the smoothed data can be problematic (Chipperfield et al., 2022)."
And in the conclusions in Line 706-708:
"In addition, the smoothed dataset should be used for quantitative analysis with great caution, as it has not been quantitatively evaluated in any way."

15. Line 79: should be "lower stratosphere". Above 25 km the lifetime of ozone becomes shorter.
Response: Revised. Thank you.

16. Line 263: should "tropospheric" not be deleted here? Otherwise, why not do stratospheric as well here?
Response: Deleted. Thank you.

17. Line 343: I don't see a comparable or better performance of MLS here, unless you mean smaller RMS / error bars, which are barely visible. In this context, see my suggestion above for Fig. 3, to add the RMS profiles to the plots, and to reduce the number of panels.
Response: Thanks for you correction. Yes, TOST has a comparable or better performance than MLS. Revised.

18. Line 390: 3d should proably be 7d
Response: Revised. Thank you.