Dear Reviewer,

Thank you for your review and for taking the time to evaluate our manuscript. We are sorry to hear your negative judgment, but we have worked diligently to address your concerns through substantial revisions. We have extensively modified the paper also in response to your comments, and we hope these changes will help convince you otherwise. Below, we have included your comments in italics, along with our responses in bold.

"General remarks:

This paper presents (i) the development of a merged ozonesonde data combining vertical ozonesonde profiles collected from the SHADOZ, NDACC and WOUDC data archives to obtain better coverage in space and time (1979-1999 and 2000-2022); (ii) the study of the impact of the sampling frequency on long term trends derived from this merged ozonesonde data set at four pressure segments (1-50 hPa, 50-100 hPa, 100-200 hPa and 200-300 hPa and five latitude bands (SP: 90-60°S; SH: 60-30°S; TR: 30°S-30 °N; NH: 30-60°N; SP: 60-90°N). Hereby, different linear statistical models have been applied to derive the trends and discussed on their representativeness and the impact of the sampling frequency on that."

We made several significant improvements to our study:

1. **We added the trend estimation of ozone concentration based on the LOTUS multiple linear regression and compared with parametric and non-parametric linear regressors. This addition allows for a more comprehensive analysis of the trends.**
2. **We calculated the representativeness of the stations at various pressure levels (10 hPa, 50 hPa and 100 hPa) using the ozone monthly gridded profiles derived from nadir-viewing satellite instrument merged in the MERGED-NP dataset, available in the Copernicus Climate Data Store (), that merges 5 vertical profiles products from UV sensors GOME, GOME2-A, GOME2-B, OMI and SCIAMACHY, still following the approach developed in Weatherhead et al. (2017). This expands the characterization of the stations clusters showing their value to characterize the trends at different latitudes and for different vertical ranges.**
3. **We excluded the analysis of trends for the South Pole (SP) and Southern Hemisphere (SH), as suggested by the editor in his review, due to the limited data availability and the insufficient number of stations, which do not guarantee reliable trend estimates.**
4. **We calculated the uncertainties of the trends for non-parametric methods and the LOTUS using the bootstrapping method. This approach provides a more robust assessment of the trend significance.**

“This merged ozonesonde dataset, as the authors assigned it as an “unified ozonesounding dataset, offering a harmonized and quality checked dataset that eliminates the necessity for climate researchers to download multiple data sets and etc….” (Line 563-566). This in itself is a big challenge to aim for and may be very helpful for the atmospheric research community. However, to achieve this goal the authors fail here in several important aspects. This so called “unified data set” is far from fulfilling the hard criteria to pretend to be a standard scientific dataset that can be used in climate research. Technically spoken, the authors have collected the sounding data from the different archives and combined them on a common grid with some statistical cleaning. However, this is neither original nor scientifically sufficient to be published in AMT.”

The primary goal of our unified dataset is indeed to provide a single, globally comprehensive database that eliminates the need for researchers to download multiple datasets. However, we also aim to study that the data produced by this unification is consistent with those already published in the literature. To achieve this, we estimated trends in ozone concentration, which is the most critical variable in this dataset, and compared them with the trends reported in the literature, using a subsampled cluster (LC cluster) compared to the entire database (LMC cluster), based on the time series of each station. Our analysis shows that the trends calculated from our unified dataset are very close to those derived from satellite datasets, both in terms of magnitude and uncertainty, although we do not pursue any homogenization effort of the time series, in the same way made for the HEGIFTOM dataset.

An additional objective of this work is to facilitate the work of data users, who often face the challenge of selecting which stations to use in order to estimate ozone variability. It is important to quantify how these station selections might impact trend values and their associated uncertainties. By providing a unified dataset with consistent quality, we aim to support users in making informed decisions and reducing uncertainty in their analyses. The unified dataset also removes a large quantity of ozonesounding profiles not properly reported, but at present available thought the data archives of the considered networks.

Ultimately, this work serves as a tool to guide users, not necessarily the most expert, encouraging them to engage with the data critically and consider the challenges associated with using discrete samples to represent continuous ozone field.

“Completely missing are scientific efforts addressing harmonization, homogenization or solid quality assessment of the ozonesonde data, although, many efforts over last two decades have been undertaken within the ozonesonde experts’ community and reported in scientific literature frequently. Many important references have not been discussed in a proper way or are completely missing in the paper. Obviously, the authors did not have consulted and evaluated the scientific literature on the performance of the different types of ozonesondes used over 4-5 decades of long term sonde records that are stored in the different data archives. This is clearly demonstrated that an important quality criterium as the total ozone column normalization factor has been completely ignored by the authors. Also in this context, the eventual impact on

the trends derived for those stations that apply the normalization factor through linear scaling of the measured vertical ozone profile. Neither the authors describe or discuss the fact that the ozonesonde community has undertaken a big homogenization effort to re-process all global ozonesonde records according the same principles (removal all known biases, consistent correction procedures, referring to same standard), including uncertainty calculation."

We acknowledge the importance of harmonization, homogenization, and quality assessment in creating a reliable dataset. The HEGIFTOM dataset has been mentioned in previous version and now more broadly described.

Furthermore, also in our response to the Editor, we clearly specified at line 103 that: "The Unified Ozonesonde Dataset merges data as the same are provided by the considered networks: this may imply the merging of higher and lower quality data (e.g. homogenized in SHADOZ vs. non-homogenized in WOUDC or NDACC) and the comparison of data from the same station provided under a different quality to different data archives, potentially affecting the bias between different data points.".

Additionally, we described the approach adopted for trend uncertainty calculation, based on the bootstrapping method. We added the following explanation to Section 3.2, at line 316, of the manuscript: "Uncertainty of the regression slopes, for the robust regressors, have been estimated using the bootstrapping method. Bootstrapping is a statistical technique that involves resampling the data with replacement to create numerous simulated samples. This method allows for the estimation of the sampling distribution of a statistic and provides measures of accuracy such as confidence intervals and standard errors (Tibshirani & Efron, 1993). By applying bootstrapping, we provide a confidence in the estimated ozone trends (Hou & Shen, 2022)."

The unified dataset has been implemented applying comprehensive data quality checks, as detailed in Section 2.4 of our manuscript. We also added a detailed explanation of how these quality checks are applied rephrasing the sentence at line 238 to provide more details: "The quality checks mentioned above are employed collectively to generate a single quality flag for each ozone profile. This flag indicates the percentage of successful data checks, with possible values ranging from 0 to 3. The plausibility and outliers' checks are used to exclude anomalous or implausible values from the database, while the remaining three checks assess the structural quality of the ozonesounding profiles. The final quality flag is determined by the number of structural checks passed, with higher values indicating better quality profiles."

Regarding the quality criterion based on the total ozone column normalization factor, we recognize that this factor is an important aspect of ozonesonde data quality. However, it could not be applied to all stations in our study due to variations in the availability and quality of total ozone column measurements across different sites.

We have included a discussion in the revised manuscript, at line 253, to explain the limitations and challenges associated with the application of the total ozone column

normalization factor: "Furthermore, we acknowledge the importance of the total ozone column normalization factor as a quality criterion for ozonesonde data. However, the application of this factor was not feasible for all stations due to variations in the availability and quality of total ozone column measurements across different sites. The normalization factor, calculated as the ratio of the spectrophotometer total ozone column (TOC) and the ozonesounding TOC, requires consistent and reliable TOC measurements (Ancellet et al., 2022). In cases where such measurements were not available or were of insufficient quality, the normalization factor could not be applied (Tarasick et al., 2021; Stauffer et al., 2022). This limitation impacts the derived trends, as the normalization factor helps to correct for biases in the measured vertical ozone profiles. Consequently, the absence of this correction for certain stations may introduce uncertainties in the trend analysis (Stauffer et al., 2022; Ancellet et al., 2022)."

Finally, we reviewed the scientific literature on the performance of different types of ozonesondes. This includes studies on the characteristics, biases, and corrections of ozonesonde data. Specifically, we added citations to the following papers, as suggested by the Editor and anonymous reviewer 1, to improve the existing knowledge base on which our study is based:

- **Hou, Y., & Shen, Z. (2022). Research Trends, Hotspots and Frontiers of Ozone Pollution from 1996 to 2021: A Review Based on a Bibliometric Visualization Analysis. *Sustainability*, *14*(17), 10898. https://doi.org/10.3390/su141710898.**
- **Tibshirani, R. J., & Efron, B. (1993). An introduction to the bootstrap. *Monographs on statistics and applied probability*, *57*(1), 1-436.**
- **World Meteorological Organization (WMO). Scientific Assessment of Ozone Depletion: 2022, GAW Report No. 278, 509 pp.; WMO: Geneva, 2022.**
- **Thompson, A. M., Stauffer, R. M.,Wargan, K., Witte, J. C., Kollonige, D.E., & Ziemke, J. R. (2021). Regional and seasonal trends in tropical ozone from SHADOZ profiles: Reference for models and satellite products. Journal of Geophysical Research: Atmospheres,126, e2021JD034691.**
- **Stauffer, R. M., Thompson, A. M., Kollonige, D. E., Komala, N., Al-Ghazali, H. K., Risdianto, D. Y., Dindang, A., Fairudz bin Jamaluddin, A., Sammathuria, M. K., Zakaria, N. B., Johnson, B. J., and Cullis, P. D.: Dynamical drivers of free-tropospheric ozone increases over equatorial Southeast Asia, Atmos. Chem. Phys., 24, 5221–5234, https://doi.org/10.5194/acp-24-5221-2024, 2024.**
- **Roeland van Malderen, Anne M Thompson, Debra E Kollonige, Ryan M Stauffer, Herman G J Smit, et al.. Global Ground-based Tropospheric Ozone Measurements: Reference Data and Individual Site Trends (2000–2022) from the TOAR-II/HEGIFTOM Project. *Atmospheric Chemistry and Physics*, 2025, ⟨10.5194/egusphere-2024-3736⟩. ⟨hal-04901618⟩**
- **Roeland van Malderen, Zhou Zang, Kai-Lan Chang, Robin Björklund, Owen R Cooper, et al. Ground-based Tropospheric Ozone Measurements: Regional**

tropospheric ozone column trends from the TOAR-II/ HEGIFTOM homogenized datasets. *Atmospheric Chemistry and Physics*, 2025, ⟨10.5194/egusphere-2024-3745⟩. ⟨hal-04901762⟩

- **Tarasick, D. W., Smit, H. G. J.,Thompson, A. M., Morris, G. A., Witte,J. C., Davies, J., et al. (2021). ImprovingECC ozonesonde data quality:Assessment of current methods andoutstanding issues. Earth and SpaceScience, 8, e2019EA000914. https://doi.org/10.1029/2019EA000914**
- **Ancellet, G., Godin-Beekmann, S., Smit, H. G. J., Stauffer, R. M., Van Malderen, R., Bodichon, R., and Pazmiño, A.: Homogenization of the Observatoire de Haute Provence electrochemical concentration cell (ECC) ozonesonde data record: comparison with lidar and satellite observations, Atmos. Meas. Tech., 15, 3105–3120, https://doi.org/10.5194/amt-15-3105-2022, 2022.**
- **Morgan MG, Henrion M. The Propagation and Analysis of Uncertainty. In: *Uncertainty: A Guide to Dealing with Uncertainty in Quantitative Risk and Policy Analysis*. Cambridge University Press; 1990:172-219.**
- **Olsen, M. A., Manney, G. L., & Liu, J. (2019). The ENSO and QBO impact onozone variability andstratosphere-troposphere exchangerelative to the subtropical jets. Journalof Geophysical Research: Atmospheres,124, 7379–7392. https://doi.org/10.1029/2019JD030435**

"At present within the HEGIFTOM activity these data have been collected and stored on an ftp-server (for details see HEGIFTOM website: https://hegiftom.meteo.be). Under the bottom line, the paper misses any ozonesonde expertise at all, and maybe the authors should have consulted or included such expertise in their study from the very beginning."

We would like to clarify that our unified database already incorporates ozonesonde data from three major networks: SHADOZ, WOUDC, and NDACC, which are also included in the HEGIFTOM dataset. While we were aware of the HEGIFTOM, the same was presented by the authors as a published dataset when the data analysis for this paper was already started in the frame of the C3S activities.

"The dilemma of sampling frequency and geographical coverage has been treated rather poorly. Almost only based on a few general statistical thresholds the different station ozonesonde records have been selected and classified in three different clusters (long, middle and short coverage resp.). However, a clear scientific rational is missing."

We have revised the manuscript to include a more detailed explanation of the criteria and their application, at lines 164-170, as follows: "Data coverage plays a critical role in accurately estimating anomalies and trends while minimizing uncertainties. To address this, only stations with at least one ozonesonde profile available per month were included in the unified dataset. These stations were then grouped according to their monthly coverage, with a month deemed covered if at least one ozonesonde ascent was available.

Based on their temporal coverage, the 153 selected stations were classified into three categories:

- **Long coverage (LC): 30 stations with a continuous data time series of at least 20 years with at most 5% of months not covered.**
- **Medium Coverage (MC): 17 stations with a continuous data time series of at least 10 years, but not more than 20 years, with at most 5% of months not covered.**
- **Short coverage (SC): 106 stations with continuous data time series less than 10 years, with at most 5% of months not covered, or no data available between 1978 and 2022."**

These thresholds were chosen to identify stations with the most reliable and extensive time series, which is key for estimating decadal trends. Each station's data was carefully investigated to improve the trend analysis.

"The discussion on representativeness through a correlation analysis of total ozone columns between a station and the neighboring points by utilizing the EP TOMS satellite data has been done rather poorly and it does not give any information on what the consequences would be for the different vertical (i.e. pressure) ranges. A more detailed correlation analysis could be obtained using other satellites deriving vertical profiles (e.g. MLS). Particularly, this would be necessary in under-sampled regions like in the tropics."

As already mentioned above, we calculated the representativeness of the stations at various pressure levels (10 hPa, 50 hPa and 100 hPa) using the ozone monthly gridded profiles derived from nadir-viewing satellite instrument merged in the MERGED-NP dataset, available in the Copernicus Climate Data Store (https://cds.climate.copernicus.eu/datasets/satellite-ozone-v1?tab=overview), that merges 5 vertical profiles products from UV sensors GOME, GOME2-A, GOME2-B, OMI and SCIAMACHY, still following the approach developed in Wheatherhead et al. (2017).

We rewrote section 2.5 as follows: "Before examining trends in the Upper Troposphere/Lower Stratosphere (UT/LS), the representativeness of the network was initially assessed to ensure an accurate capture of ozone variability across different latitudes. This assessment utilized an approach developed by Weatherhead et al. (2017), utilizing the MERGED-NP dataset, that merges 5 ozone monthly gridded vertical profiles products, derived from nadir-viewing satellite instrument, from UV sensors GOME, GOME2-A, GOME2-B, OMI and SCIAMACHY. The dataset is available in the Copernicus Climate Data Store (https://cds.climate.copernicus.eu/datasets/satellite-ozone-v1?tab=overview) and provides measurements from January 2003 to December 2020. The correlations were calculated for deseasonalized monthly values. Spatial representativeness of the LMC and LC datasets for ozone concentrations was quantified using Pearson correlation coefficient for each grid point with each unified station considering specific vertical ranges (170 hPa, 100 hPa, 50 hPa and 1 hPa) provided by

MERGED-NP dataset, as depicted in Figure 9. According to Weatherhead et al. (2017), a good correlation is indicated by a coefficient higher than 0.7.”

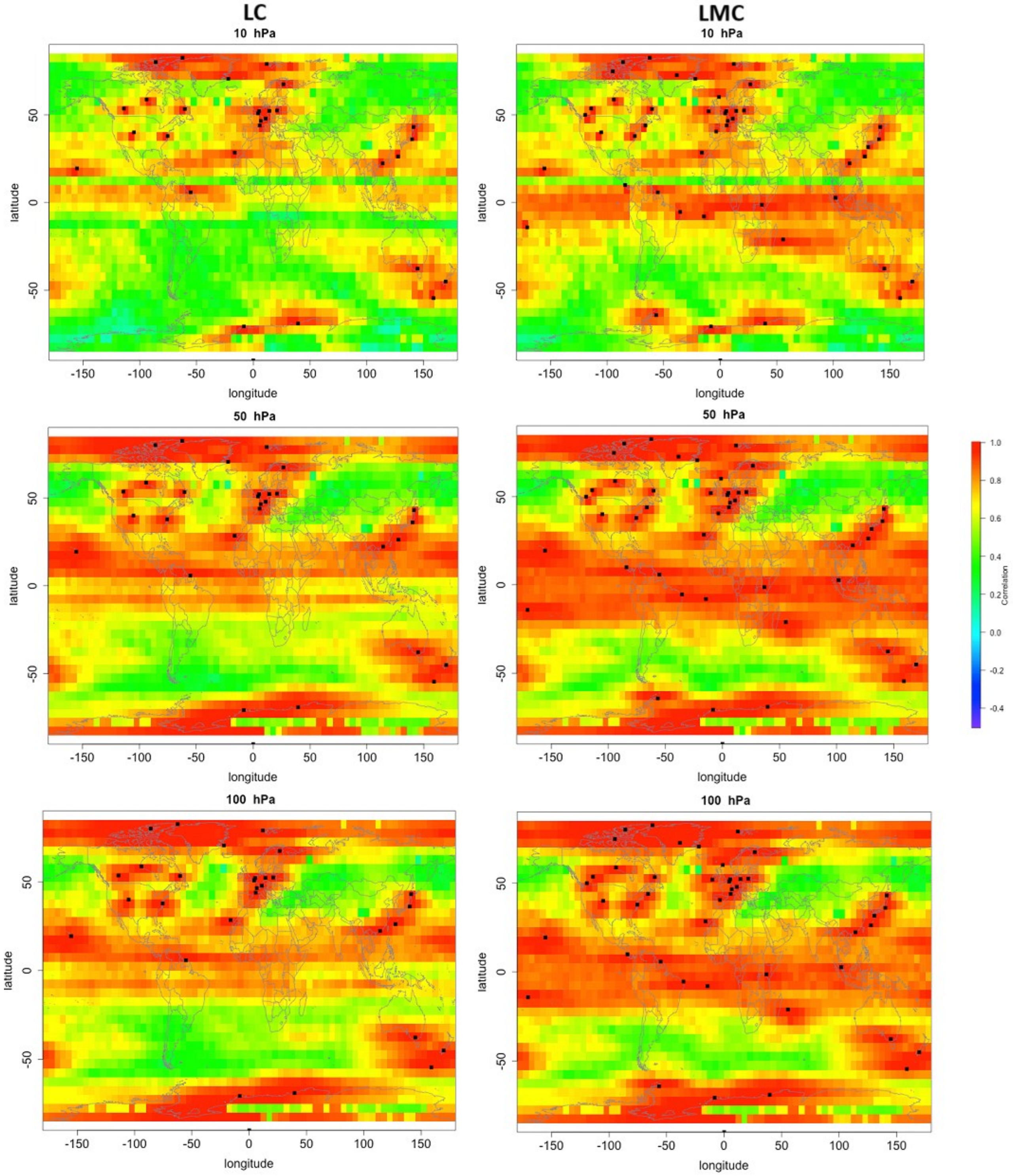

Figure 9. Assessment of representativeness for unified stations, with the left column depicting the LC cluster at 10 hPa (upper panel), 50 hPa (central panel) and 100 hPa (lower panel) and the same for the right panel showing the LMC cluster, utilizing the MERGED-NP dataset from January 2003 to December 2020. The map illustrates the correlation level between each station and its surrounding area. The linear Pearson correlation coefficients were calculated between the deseasonalized monthly averages from the ozone profile measurements. A correlation value of 1 indicates a complete correlation with the surrounding grid points, while 0 signifies no correlation with neighbouring points. A good correlation is indicated by a coefficient higher than 0.7.

Figure 9 reveals the stations' representativeness for the LC (left column) and LMC (right column) at three different pressure levels: 10 hPa (upper row), 50 hPa (central row) and 100 hPa (lower row). These levels were considered the most representative among those available in the MERGED-NP dataset (1 hPa, 5 hPa, 10 hPa, 50 hPa, 100 hPa, 170 hPa and 450 hPa), with respect to the vertical ranges chosen for trend estimation. Figure 9 highlights that the LMC cluster generally exhibits greater representativeness than the LC cluster, particularly in the TR and the SH. This difference is particularly evident at 10 hPa, where LC stations show significantly lower representativeness compared to LMC stations, especially in the TR. At 50 hPa and 100 hPa, the representativeness of both clusters is largely similar, with only minor variations. In the other latitudinal sectors, the difference between the LC and LMC clusters is not as evident as in the TR and SH sectors, resulting in the LC cluster having a similar representativeness to the LMC cluster in those regions.

Consequently, the additional stations within the LMC cluster may be deemed redundant for total column and stratospheric ozone for these regions, with preference given to utilizing data from the LC station due to its higher-quality correlated time series. However, with different station launch schedules, this density remains important and not redundant for tropospheric measurements. Furthermore, it is worth noting that, at NH, the correlation is strong for only a few regions (North America, Europe and Japan), as the LC and LMC clusters do not include stations in the other regions. Given the limited representativeness of the LMC and LC datasets for the entire total ozone band of NH, it is difficult to draw definitive conclusions about trends in this region. This limitation implies that NH trends calculated from these datasets may not fully capture the ozone variability in all areas within this latitude band. Consequently, trends observed in different vertical ranges should be interpreted with caution, recognizing the potential gaps in spatial coverage and representativeness.

Figure 9 also reveals that the representativeness of ozone at 10 hPa is lower compared to that at 50 hPa and 100 hPa, primarily due to greater spatial and temporal variability in the upper stratosphere. This variability is influenced by complex dynamic and chemical processes, such as the Brewer-Dobson circulation and photochemical reactions, which cause increased ozone variability. Additionally, the density of measurements at 10 hPa may be lower due to instrumental limitations, reducing the representativeness of stations to the surrounding areas. Planetary waves, which significantly impact the distribution of ozone in the upper stratosphere, can also contribute to reducing the correlation between station measurements and surrounding areas (WMO, 2022). However, despite these challenges, the representativeness at 10 hPa is considered acceptable, as correlation coefficients are above 0.7 for most of the sector, indicating good correlation with the surrounding grid points."

"Further, all trends should have been reported with their uncertainties, even when the trend is little and should be not labeled with "NT"."

As also reported in the replies to Reviewer 1 and the Editor, we have modified Tables 9-12 to include trend uncertainties, calculated via bootstrapping, for each regressor. Additionally, we have incorporated the LOTUS regressor along with the methods used previously. We have focused on reporting trends for the Northern Hemisphere (NH), the Tropics (TR), and the North Pole (NP), as data availability and the number of stations for the Southern Hemisphere (SH) and the South Pole (SP) are insufficient to ensure reliable trend estimates. These modifications aim to provide a clearer understanding of both significant and non-significant trends. Below, as also reported in our response to the editor's review, are the new Tables 9-12 with updated discussion at lines 391-442:

"

Table 9. Trend estimates, with the uncertainty, as percentage per decade (% dec-1) obtained with the Linear (LIN), Least Absolute Deviation (LAD) and Theil-Sen (TS) and LOTUS regressors for all latitudinal sectors considered at the 50-1hPa vertical range for each cluster. Legend: Positive value (increasing trend); Negative value (decreasing trend); NT (No Trend).

50-1hPa		% dec^{-1}	TR	NH	NP
1978-1999	LC	LIN	7 ± 1	-5.9 ± 0.5	-10 ± 2
		LAD	8 ± 1	-5 ± 1	-10 ± 6
		TS	7 ± 2	-6 ± 2	-8 ± 8
		LOTUS	8 ± 1	-6.3 ± 0.2	-10 ± 1
	LMC	LIN	4 ± 1	-5.6 ± 0.5	-14 ± 2
		LAD	5 ± 2	-5 ± 1	-12 ± 3
		TS	7 ± 3	-6 ± 2	-11 ± 5
		LOTUS	4 ± 1	-6.0 ± 0.2	-14.9 ± 0.4
2000-2022	LC	LIN	NT	-1.1 ± 0.4	-6 ± 1
		LAD	NT	-1 ± 1	-5 ± 2
		TS	NT	-1 ± 1	-6 ± 2
		LOTUS	NT	-1.1 ± 0.1	-5.4 ± 0.2
	LMC	LIN	0.7 ± 0.2	NT	-5 ± 1
		LAD	0.6 ± 0.6	NT	-5 ± 2
		TS	0.7 ± 0.7	NT	-5 ± 2
		LOTUS	0.9 ± 0.1	NT	-4.6 ± 0.2

At the 50-1 hPa vertical range (Table 9) for the TR during the period 1978-1999, estimates display discrepancies of up to about 4% ±2% per decade. This discrepancy likely arises from the greater number of stations available for the LMC cluster, enhancing coverage and representativeness compared to the LC cluster, as discussed in Section 2.5.

For the NH, the clusters agree for the 1978-1999 period (with a maximum discrepancy less than 0.5% ±0.7% per decade), but in contrast for 2000-2022, where the trend of the LMC cluster, for the MK test, is not significant.

Finally, in the NP, trends correspond only in the 2000-2022 period (with a maximum discrepancy of 1% ±2% per decade for the TS regressor). For 1978-1999, differences among clusters nearly reach 5% ±2% per decade.

Table 10. Same as Table 9 but for the vertical range 100-50 hPa.

100-50hPa		% dec^{-1}	TR	NH	NP
1978-1999	LC	LIN	12 ± 2	-9 ± 1	NT
		LAD	14 ± 7	-8 ± 1	NT
		TS	11 ± 10	-9 ± 2	NT
		LOTUS	11 ± 1	-8.7 ± 0.4	NT
	LMC	LIN	8 ± 2	-10 ± 1	-13 ± 1
		LAD	8 ± 7	-10 ± 2	-12 ± 3
		TS	2 ± 9	-9 ± 3	-12 ± 4
		LOTUS	8 ± 1	-10.4 ± 0.4	-12.9 ± 0.5
2000-2022	LC	LIN	4 ± 1	3 ± 1	NT
		LAD	3 ± 2	4 ± 1	NT
		TS	4 ± 3	4 ± 1	NT
		LOTUS	2.5 ± 0.2	2.8 ± 0.1	NT
	LMC	LIN	4 ± 1	3 ± 1	NT
		LAD	3 ± 2	3 ± 1	NT
		TS	2 ± 3	3 ± 1	NT
		LOTUS	1.6 ± 0.4	2.9 ± 0.1	NT

In Table 10, focusing on the 100-50 hPa range, the NH sector stands out with trend estimates in the period 2000-2022 showing 1% ±1% per decade sampling error and significant trend estimates, attributed to higher data density and greater variability in this vertical range. Nonetheless, even in NH, the LC and LMC clusters exhibit discrepancies in the period 1978-1999 of around 2% ±3% per decade in estimated trends. Conversely, the NP sector lacks significant trends, except for the LMC cluster during 1978-1999, with especially noticeable trends.

Additionally, the TR sector presents a considerable discrepancy, in the 1978-1999 period, of 9% ±10% per decade while, in the period 2000-2022, the LC and LMC clusters show

discrepancies of 2% ±3% per decade. Interestingly, the trends estimated by the LMC and LC clusters show low discrepancies among regressors with structural uncertainties of 2.4% ±3% per decade and 1.5% ±3% per decade for 1978-1999 respectively. The same cannot be said for 2000-2022 due to the discrepancies of 6% ±9% per decade for the LMC cluster and 3% ±10% per decade for the LC. This performance can be attributed to the paucity of available data for the TR, in 1978-1999, despite the LMC cluster comprised 8 more stations than LC (see Table 2). For conducting trend analysis using data from the unified database, high-quality measurements similar to those from the LC cluster are crucial. However, in regions with a limited number of LC stations, such as the tropics (TR), data from the LMC cluster can offer better representativeness than the LC cluster.

Table 11. Same as Table 9 but for the vertical range 200-100 hPa.

200-100hPa		% dec^{-1}	TR	NH	NP
1978-1999	LC	LIN	-	-10 ± 1	NT
		LAD	-	-10 ± 3	NT
		TS	-	-10 ± 5	NT
		LOTUS	-	-10 ± 1	NT
	LMC	LIN	10 ± 3	-13 ± 1	-19 ± 2
		LAD	8 ± 9	-14 ± 3	-15 ± 4
		TS	10 ± 15	-13 ± 5	-19 ± 6
		LOTUS	10 ± 1	-13 ± 1	-19 ± 1
2000-2022	LC	LIN	10 ± 1	7 ± 1	NT
		LAD	9 ± 4	7 ± 2	NT
		TS	8 ± 4	5 ± 3	NT
		LOTUS	7.5 ± 0.4	8.1 ± 0.2	NT
	LMC	LIN	6 ± 1	6 ± 1	NT
		LAD	3 ± 4	6 ± 2	NT
		TS	6 ± 4	8 ± 3	NT
		LOTUS	2.3 ± 0.4	7.5 ± 0.2	NT

In Table 11, trend estimates in the NP sector are non-significant except for the LMC cluster in 1978-1999. Trend assessments for the LC cluster in the TR for the period 1978-1999 should be approached with caution due to the significantly small amount of data available before 1995 compared to subsequent years, which largely inflates uncertainties on decadal trends. However, it is important to acknowledge that the absence of a significant trend could indeed be a valid physical observation, particularly in regions or periods where ozone dynamics are more stable or less prone to significant change.

In addition, for 2000-2022, the LC cluster reveals a structural uncertainty of 2.5% ±4% per decade, and a discrepancy with the LMC of 6% ±4% per decade. Conversely, trend estimates for the LMC cluster in TR are significant, but exhibit considerable discrepancies among regressors, with structural uncertainty of 2% ±15% per decade for 1978-1999 and 3.7% ±4% per decade for 2000-2022. Moreover, the NH sector is the only one with all valid trends, although there is a disagreement between LC and LMC (up to about 4% ±5% per decade for 1978-1999 and 3% ±3% per decade for 2000-2022). Furthermore, it is worth considering the discrepancy between the results of the TS regressor with those of the LIN and LAD regressors (2% ±3% and 3% ±4%, respectively) that occurred for LC in the period 2000-2022. The LIN, LAD and LOTUS regressors agree with each other, reporting a structural uncertainty of 1.1% ±1% per decade, on the contrary TS where this uncertainty rises to around 3% ±3% per decade.

Table 12. Same as Table 9 but for the vertical range 300-200hPa.

300-200hPa		% dec^{-1}	TR	NH	NP
1978-1999	LC	LIN	17 ± 3	-10 ± 1	5 ± 5
		LAD	19 ± 7	-10 ± 4	11 ± 11
		TS	17 ± 11	-13 ± 6	10 ± 13
		LOTUS	17 ± 1	-10.2 ± 0.5	5 ± 2
	LMC	LIN	NT	-12 ± 2	-
		LAD	NT	-11 ± 3	-
		TS	NT	-11 ± 6	-
		LOTUS	NT	-12.8 ± 0.4	-
2000-2022	LC	LIN	5 ± 1	6 ± 1	NT
		LAD	3 ± 3	7 ± 2	NT
		TS	4 ± 3	7 ± 4	NT
		LOTUS	3.7 ± 0.2	7.6 ± 0.4	NT
	LMC	LIN	NT	6 ± 1	NT
		LAD	NT	7 ± 2	NT
		TS	NT	5 ± 3	NT
		LOTUS	NT	7.2 ± 0.4	NT

The scenario for the 300-200 hPa layer (Table 12) is similar to the previous layer. Additionally, this vertical interval exhibits the fewest significant trends. The most suitable vertical ranges for trend analysis are 50-1 hPa and 100-50 hPa, due to their more stable ozone concentrations and lower variability. In contrast, the high variability of ozone near the tropopause complicates the detection of trends in that region. Table 13 highlights the

most and least significant disparities in the trends obtained by comparing the different non-parametric regression techniques used with the parametric regressor LOTUS, taken as a reference, facilitating the discussion of the structural uncertainties. The LOTUS regressor was taken as a reference because of the use of indicators such as ENSO (El Niño-Southern Oscillation) and QBO (Quasi-Biennial Oscillation) within it that allow a refinement of the trend estimates. These indicators can have variable impacts depending on the latitude band, improving the accuracy of the trend analysis in some regions and reducing the influence in others (Olsen et al., 2019). In comparison between the regressors, the uncertainties were propagated in quadrature (Morgan & Henrion,1990; Stauffer et al., 2022)."

"Summarizing, in the present form the paper misses almost any scientific originality. Without having a solid quality assessment of the merged ozonesonde dataset and their representativeness, the paper misses the scientific base to be published in AMT, therefore, I rate the paper as scientifically poor and reject it for publication in AMT."

We have taken the reviewer's comments very seriously and made substantial revisions to address the concerns raised. We hope that the revisions added to the paper demonstrate the scientific utility of this work for the ozone sounding data users, and we kindly request a reconsideration of the reviewer's assessment based on the improvements made.