Dear Reviewer,

Thank you very much for your thorough review and for the time you have dedicated to evaluating our paper. We greatly appreciate your positive feedback and constructive comments. Below, we have included your comments in *italics* and our responses in bold.

"Review of "Effect of sampling error on ozone partial pressure trends within a unified ozonesounding dataset" Fabrizio Marra, et al.

Summary and General Comments:

This paper describes the construction of a "unified" ozonesonde dataset gathered from WOUDC, SHADOZ, and NDACC archives and the calculation of ozone profile trends in segments from the upper troposphere to the lower stratosphere for several latitudinal bands and two time periods. Various methods are used to derive the trend estimates to understand their uncertainties and how data sampling affects them. The authors find that ozone trends calculated over the periods 1978-1999 and 2000-2022 from their unified dataset agree fairly well with independent studies that include satellite datasets with greater spatiotemporal sampling.

This paper is well motivated and written. Analyzing a fully global set of ozonesonde data that appears in various archives is a challenge, so this study results in two key accomplishments: 1) constructing a quality-checked global ozonesonde dataset that avoids duplication and captures data that would be missed by using only one archive, and 2) demonstrating the utility of the unified dataset for trends calculations and comparison to existing studies."

Thank you for your positive feedback and for acknowledging the key findings of our study. We are glad to hear that you found our article well motivated and written.

"Uncertainties of the trends (95% confidence interval at least, and perhaps also p-value) presented in Tables 9-12 should be reported so we can understand the significance (or insignificance) of the differences in trends computed using various methods. This would also allow you to also report a trend that is not statistically significant, rather than listing "NT"."

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)."

Following the reference added:

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

"The fact that trends computed using the unified ozonesonde dataset generally agree, except for 1978-1999 tropical stratospheric trends, with independent studies that include satellite data is an important message. This should be communicated in the abstract and will lead to the paper having greater impact."

We have revised the abstract to emphasize the alignment of trends derived from the unified ozonesonde dataset with independent studies that incorporate satellite data. Below is the revised abstract: **"This work discusses the impact of the sampling frequency on ozone partial pressure trends, estimating its impact at various latitudes and vertical layers in the upper troposphere/lower stratosphere (UT/LS) region. The trends are estimated in the periods 1978-1999 and 2000-2022, using a new unified dataset combining the ozonesounding profiles provided by SHADOZ (Southern Hemisphere ADditional OZonesondes), NDACC (Network for the Detection of Atmospheric Composition Change), and WOUDC (World Ozone and Ultraviolet Radiation Data Centre). These datasets are combined to offer adequate coverage at various latitudes and to enhance the estimation of anomalies and trends in ozone concentration on a global scale. The available measurements are classified into three groups (short coverage, medium coverage and long coverage) based on the temporal coverage of historical time series. The representativeness of medium coverage and long coverage cluster have been also studied at each altitude level using independent nadir-viewing satellite data. Parametric, non-parametric and multiple linear regressors are utilized to estimate trends and the related differences to quantify structural uncertainty. Significant trends for the period 1978-1999 are estimated for the Northern Hemisphere mid-latitude (NH), which shows a negative trend of 6% ±1% per decade in the layer 50-1 hPa and a negative trend of 9% ±1% per decade at 100-50 hPa, and for the Tropics (TR), which shows a positive trend of about 5% ±2% per decade at 50-1 hPa and 8% ±2% per decade at 100-50 hPa, respectively. Regarding the 2000-2022, the NH reveals a negative trend of 1% ±1% per decade at 50-1 hPa and a positive trend of 4% ±1% per decade at 100-50 hPa, and the TR shows a positive trend of 0.7 ±0.6% per decade at 50-1 hPa and a positive trend of 3% ±2% per decade at 100-50 hPa. Furthermore, the sampling error between the clusters was investigated, revealing a small**

effect of less than 2% per decade at 100-50 hPa and 0.5% per decade at 50-1 hPa for NH and about 9% per decade at 100-50 hPa and 4% per decade at 50-1 hPa for TR, as well as the structural uncertainty between the regressors used, 1.2% per decade at 100-50 hPa and 1.3% per decade at 50-1 hPa for NH and 6% per decade at 100-50 hPa and lower than 3% per decade at 50-1 hPa for TR. The trends computed using the unified ozonesonde dataset generally agree with independent studies based on more sophisticated datasets, including satellite data. However, discrepancies are observed in the tropical stratospheric trends during the period 1978-1999, primarily due to the limited availability of observations."

"Recommendation:

The general and specific comments that I have should result in mostly minor corrections and additions. After they are addressed, I recommend that this paper be published.

Line-by-Line and Technical Comments:

Abstract Line 15: Please list the three groups here."

Ok, please see our previous comment regarding the Abstract.

"Line 34: Suggest also citing Thompson et al., (2021; https://agupubs.onlinelibrary.wiley.com/doi/full/10.1029/2021JD034691) and Stauffer et al., (2024; https://acp.copernicus.org/articles/24/5221/2024/). Additionally, two HEGIFTOM papers (Van Malderen et al., 2025a, b) will be posted as preprints to the TOAR-II special issue not long after this review is submitted. Those will also be excellent and contemporary references."

OK, we will add these citations into our manuscript.

"Line 36: Change "utilized" to "applied"."

OK.

"Line 36: Change "alongside" to "to"."

OK.

"Line 46: To my knowledge there are only three stations that launch ~3 times per week: Uccle, Payerne, and Hohenpeissenberg. Weekly launches are much more typical."

We rephrased line 46 as: "However, unlike regular radiosoundings, ozonesonde measurements are performed less frequently, with profiles collected typically once a week."

"Line 95: In the case of a dataset existing in more than one archive, what is done to ensure that the homogenized ozonesonde data are being used? Are there cases where duplicate profiles were found in different archives with slightly different (e.g., non-homogenized vs. homogenized) ozone values?"

To explain clearly the concept reported at lines 91-93 ("Duplications often arise when observations from the same station are transmitted to multiple networks, leading to discrepancies in transmission periods, data formats, and the number of data points, along with varying metadata. Additionally, not all networks provide identical number of data points at each level."), we added a phrase at line 103, specifying what is written in your comment: "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 vertical profiles (e.g. homogenized in SHADOZ vs. non-homogenized in WOUDC or NDACC) and the investigation of time series comprising different quality data from different data archives. The merging may potentially affect the bias between different profiles."

Regarding the cases where duplicate profiles were found in different archives with slightly different ozone values, two approaches are possible:

- **Use samples from discarded profiles to densify the vertical sampling if they provide measurements at different pressure levels not available in the selected profiles.**
- **Choose the profile with the highest number of successful quality checks, discarding the other profiles.**

The second approach was adopted as it can be considered more suitable and less prone to issues affecting one of the merged datasets.

The author used the first approach for his PhD thesis, available at https://hdl.handle.net/11563/177616, to study and compare the results with the literature by merging the missed samples from the discarded profiles.

"Figures 4, 5, and 8: Please flip the axes so pressure is the y-axis"

Ok. We modified these Figures, flipping the axes and reporting the ozone partial pressure in mPa, as follows:

Figure 4:

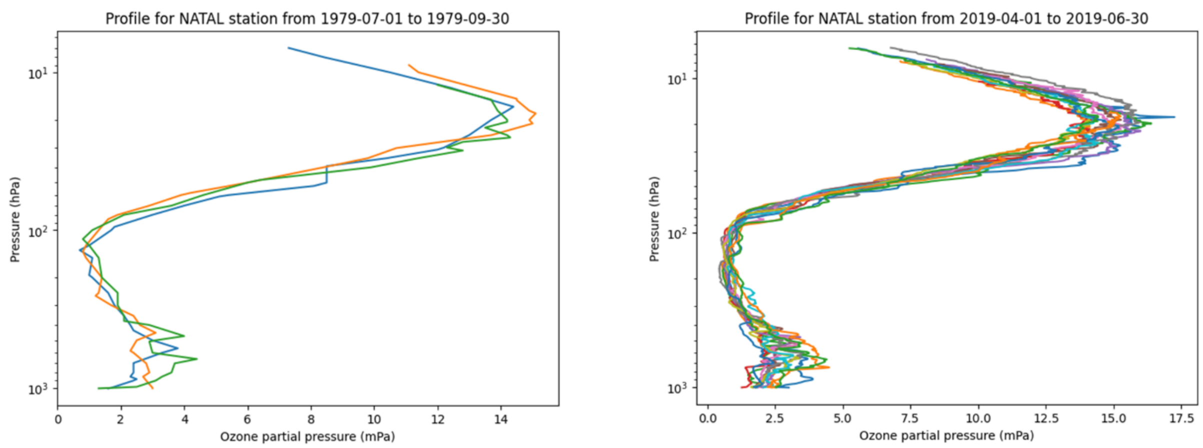


Figure 5:

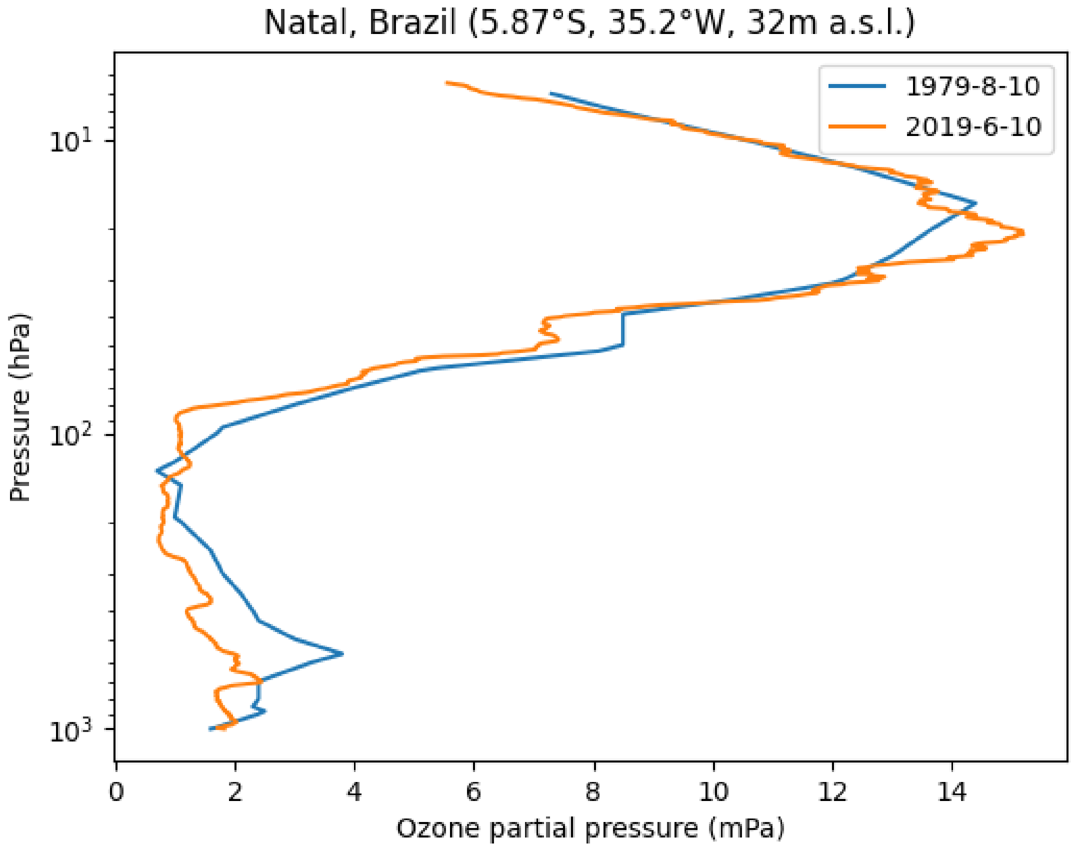
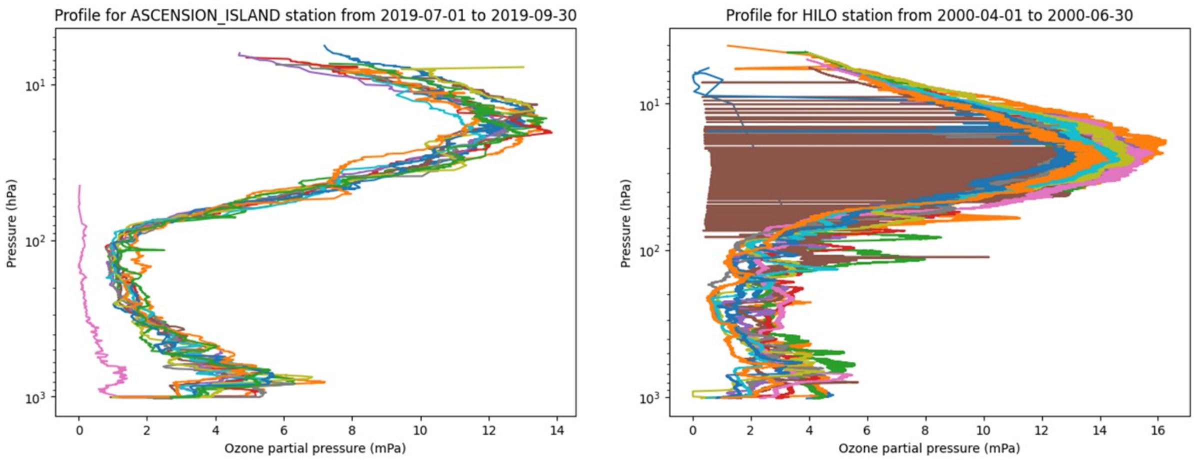


Figure 8:

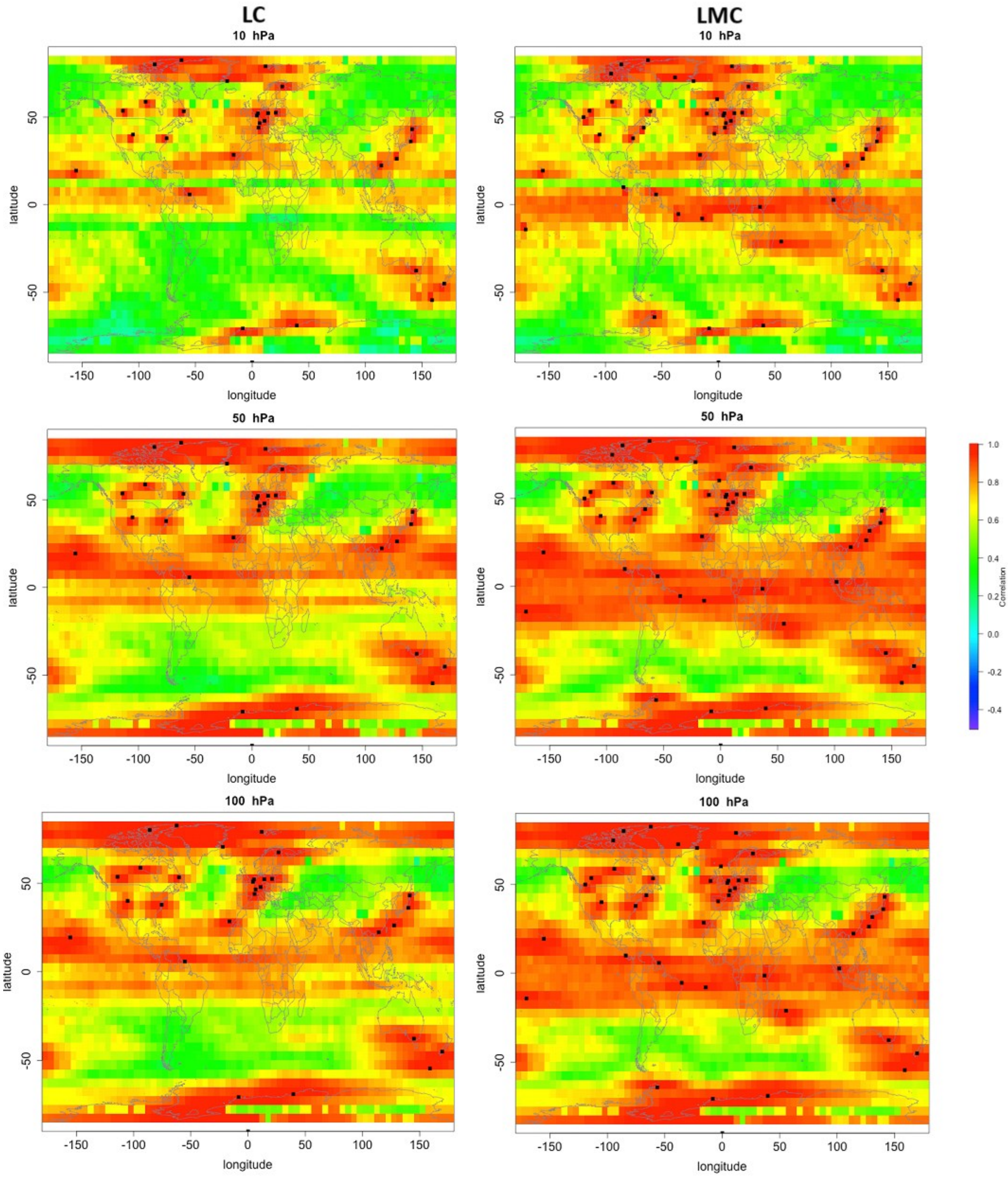


"Line 200: How did you determine not to include SP and SH even though the trends are significant? What criteria were used to exclude them? Here is where the reporting of uncertainties might help."

Regarding SH, as shown in Figure 9, the structural uncertainties (between the considered regressors) vary in a large range between 6% and 104% for 1978-1999, while for 2000-2022 the variability is smaller and comparable to that of NH values only for 50-1 hPa and 100-50

hPa pressure levels. Therefore, lines 201-205, have been rephrased as: "Due to the limited availability of LC and MC stations in the Southern Polar (SP) and Southern Hemisphere (SH) regions, the analysis presented hereinafter is not conducted on these two regions. Although the related estimated trends are significant, the paucity of available stations' number and data in these sectors substantially inflates the uncertainties on the estimated trends. Specifically, for SH, as shown in Figure 9, the number of stations stored in the database is not sufficiently representative. For the SP, despite the good representativeness of the available stations, the structural uncertainties (between the considered regressors) vary in a large range between 6% and 104% for 1978-1999, while for 2000-2022 the variability is smaller and comparable to that of NH values only for 50-1 hPa and 100-50 hPa pressure levels. Therefore, the comparative analysis is limited by the broad differences in the estimated trends among the considered regressors, suggesting that deceptive conclusions should be avoided for this latitude sector."

Below the updated Figure 9, where the representativeness of the different stations clusters is shown at different vertical ranges, 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).



"Line 219: What is meant by "5 ascents per month"? Is this for latitudinally-aggregated stations, or do you require an individual site to have at least 5 profiles in a given month over all years of the timeseries?"

We meant that an individual site has at least 5 profiles in a given month over all years of the time series. We rephrase line 219 as: "Completeness checks were performed to ensure that each individual site had at least 5 ascents per month in all years of the time

series, allowing for up to 5% of months not covered, as mentioned in the classification criteria;"

"Line 230: I am assuming the large majority of profiles that did not pass this check were from before 1990 or so, correct?"

The profiles that did not pass this check are 5439 before 1990 (29.75% of profiles that did not pass the check). In addition, considering that the total stored profiles in the database before 1990 is 12506, the percentage of these profiles that do not pass the check is 43.49%. So, asking the question, we can say that the large majority of profiles that did not pass this check were from before 1990. We rephrased line 230 as: "In this work, 79603 profiles were checked and only 18358 (23.06%) did not pass the check in certain atmospheric regions. The majority of profiles that did not pass the checks were from before 1990 (5439 profiles, 29.75%)."

"Table 4: Are these null values what you have removed based on previous quality checks? Please explain more what this means."

Not exactly: some of null values were flagged by QCs and others are not provided in the source files, e.g. uncertainty values are always provided by SHADOZ and, for a fraction of data, by NDACC. Instead, they are not available from WOUDC. Better caption, at line 236, may be: "List of the main variables within the unified database on which the percentage of null values is calculated. The latter are due to either flagging by applied QCs or because missing in the source file. For each variable, motivation for retrieving null values is reported."

"Line 269: One might argue that some closely located ozonesonde stations could be somewhat redundant for total column and stratospheric ozone. However, with differing launch schedules at the stations this density of stations remains important, and they are certainly not redundant for tropospheric measurements. Please make this clear in the paper."

Ok, we rephrased line 269 as follows: "Consequently, the additional stations within the LMC cluster may be deemed redundant for total column and stratospheric ozone in these regions, with preference given to utilizing data from the LC stations due to their higher-quality and correlated time series. However, redundant measurements must be considered relevant for tropospheric measurements."

"Section 3.2: Can you explain in more detail how the anomaly timeseries are constructed? Are the ozone values in a latitudinal band first averaged to construct a climatology, and then the anomalies are computed? Have you explored calculating anomalies for individual stations from their individual climatology, and then averaging the individual site anomalies within the latitudinal bands? This could help avoid some step changes resulting from differing station records – having stations come online, dropping out, etc."

Regarding the first two questions, we rephrased lines 306-307 to be clearer as follows: "The average monthly anomalies are calculated using data from all stations within each latitudinal and vertical range, using the following formula:".

Regarding the question "Have you explored calculating anomalies for individual stations from their individual climatology and then averaging the individual site anomalies within the latitudinal bands?": The preference was to aggregate the data at the latitude level to avoid large uncertainties in trend estimates due to gaps in many individual series. This approach may ensure more reliable trend calculations by minimizing the impact of missing data at individual stations. However, in the future, to compare this method with the one used in this work.

"Line 374: Change "As" to "For"."

OK.

"Table 9: The large ozone increases in the tropical stratosphere for 1978-1999 are really surprising, even with limited sampling, and based on Figure 12 they appear to be driven by the years ~1982-1988. Do you know what is happening here? Again, uncertainty reporting will help here."

Probably, this effect is due to the eruption of El Chichón in 1982, as I mentioned in Section 4.2, lines 521-533.

"Line 400: Change "Chapter" to "Section"."

OK.

"Line 414-415: Better to say "significant" rather than "valid"."

OK.

"Line 421: Change "an enhanced" to "more"."

OK.

"Line 427: Change "vanishing" to "inhibiting"."

OK.

"Line 440: I wouldn't use "richer data content" to describe the difficulty in detecting trends in the upper troposphere vs. the stratosphere. The high variability of ozone near the tropopause is mostly what makes it difficult to detect trends in that region."

We rephrased lines 440-441 as follows: "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."

“Line 512: Change how the ozonesonde TCO dropoff is described. Rather than stating “apparent anomalous losses of ozone in the lower and middle stratosphere”, note that it is an instrumental artefact or artefacts that cause low-biased ozone measurements in the stratosphere at some stations.”

We rephrased lines 510-514 as: “Figure 12 also shows the sudden post-2013 total column ozone (TCO) “dropoff” of a few percent in the ozone anomalies due to one of the two instruments used for the ozone soundings (Stauffer et al., 2020; Stauffer et al., 2022; Nakano and Morofuji, 2023). This dropoff is attributed to instrumental artefacts that cause low-biased ozone measurements in the stratosphere at some stations, resulting in several percentage points deviation from the averages observed from 2014 to 2017.”

“Line 568: It is just HEGIFTOM, not HEGIFTOM-II”

OK.

Appendix A: I note that for the Costa Rica station that the three closely located sites where that station has moved around to have been separated. Were these data taken from the SHADOZ archive? It is generally considered a single station, especially for measurements in the upper troposphere and stratosphere as used here.

The data for Costa Rica station are retrieved from SHADOZ and WOUDC. SHADOZ provides the data from Alajuela, Heredia and San Pedro considering them as a single station, WOUDC instead provides them in the form of three different stations. In the analysis, we only used the San Pedro data from WOUDC, classifying Alajuela and Heredia as short coverage.

In the updated manuscript, considering all the contributions from WOUDC dataset related to the Costa Rica station appropriately as from the same station, the data starts in July 2005 and ends in December 2020, with a record of 15 years and 6 months, compared to the 11 years previously obtained. All other related parts, text and calculations, in the manuscript have been adjusted accordingly.