Response to Referee #1 comment on "Ground-based Tropospheric Ozone Measurements: Regional tropospheric ozone column trends from the TOAR-II/ HEGIFTOM homogenized datasets" by Van Malderen et al.

The manuscript has presented the regional mean tropospheric ozone trends using the currently available homogenized ground-based and in-situ tropospheric ozone measurements, as a research outcome from the TOAR-II HEGIFTOM working group. Several statistical methods, including spatial correlation analysis, quantile regression, and linear mixed-effects modeling, are applied to test the consistency of tropospheric ozone trends as estimated from different regions, periods and datasets. Overall, the study is a valuable addition to our current understanding by summarizing the free tropospheric ozone trends revealed by HEGIFTOM datasets. Such datasets and estimated trends will provide benchmarks for model evaluation.

Thank you very much for your positive feedback and for taking your time to review the manuscript. Your suggestions have been taken into account carefully and we believed they improved the manuscript substantially.

Here I have several comments that the authors should address to further improve the manuscript.

Specific Comments

1) Page 2, Line 58-59, Abstract –

Here, median confidence and high confidence were used to describe the robustness of the trends; however, how were these terms defined in the main text? Some explanation was mentioned in Table 5, yet I think some more details are needed to clarify their statistical meaning.

You are right, the explanation of this concept deserves to be in the main text as well. As a matter of fact, the use of terms such as median confidence and high confidence can be traced back to IPCC AR5, as described by Mastrandrea et al. (2010, 2011). So, we are just following well-known IPCC methods and we refer the reader to these papers. In the beginning of Section 5.1.2, we added:

"Table 5 lists the trend estimates, 2-sigma uncertainties, *p*-values, and trend confidence. We hereby use the uncertainty scale for assessing the reliability and likelihood of the estimated trend that has been developed within TOAR-II (Chang et al., 2023b): very high certainty ($p \leq 0.01$), high certainty ($0.05 \geq p > 0.01$), medium certainty ($0.10 \geq p > 0.05$), low certainty ($0.33 \geq p > 0.10$), and very low certainty or no evidence ($p > 0.33$). Following the methodology of the IPCC Assessment Report 5 (Mastrandrea et al. (2010, 2011), we adopt the concept of trend confidence, which combines this trend uncertainty (based on the *p*-value and the 95% confidence interval) with the data coverage, as recently applied by Gaudel et al. (2024), based on the number of observations per month and continuity of sampling (see Table 2 for the classification). For example, higher confidence trends will be obtained for trends with lower *p*-values and higher data coverage. Table A1 in Gaudel et al. (2024) provides the translation table or calibrated language for assigning the confidence level based on trend uncertainty and data coverage, and has been applied here."

Mastrandrea, M. D., Field, C. B., Stocker, T. F., Edenhofer, O., Ebi, K. L., Frame, D. J., Held, H., Kriegler, E., Mach, K. J., Matschoss, P. R., and Plattner, G. K.: Guidance note for lead authors of the IPCC Fifth Assessment Report on consistent treatment of uncertainties, Intergovernmental Panel on Climate Change, https://www.ipcc.ch/site/assets/uploads/2017/08/AR5_Uncertainty_Guidance_Note.pdf (last access: 13 May 2025), 2010.

Mastrandrea, M.D., Mach, K.J., Plattner, G. K., Edenhofer, O., Stocker, T. F., Field, C. B., Ebi, K. L., and Matschoss, P. R.: The IPCC AR5 guidance note on consistent treatment of uncertainties: a common approach across the working groups. Climatic Change 108, 675, https://doi.org/10.1007/s10584-011-0178-6, 2011.

2) Page 7, Line 225 -

Twenty-four well-correlated regions were identified here using the spatial correlation analysis; however, it was unclear how these 24 regions were located and their covered areas. Such information was unclear by looking at Figure 1 as an example.

In Gaudel et al. (2020) and Wang et al. (2022), see their Fig. 1, 11 different regions are proposed in which IAGOS and ozonesonde tropospheric ozone profile data are merged based on co-location. For the airports and ground-based sites located within those regions, we performed the correlation analysis between the CAMS TrOC/FTOC time series at those locations. We then refined (i.e. split up in most cases) those original regions so that the correlation coefficients are higher than 0.7 between all HEGIFTOM sites within a region. As a consequence, each region contains only well-correlated sites in terms of TrOC/FTOC. By adding some sites with different co-located techniques (Mauna Loa and Lauder) and the Canadian Arctic to the original regional selection in Gaudel et al. (2020) and Wang et al. (2022), we end up with 24 well-correlated regions that are covered with ground-based HEGIFTOM data.

We tried to clarify this better in the paper. First, at the end of the first paragraph of Sect. 3.1, we changed the sentence into "We use the 11 regions defined in Gaudel et al. (2020) and Wang et al. (2022) on the basis of IAGOS and ozonesonde profile measurements as the starting points for merging time series, but the final regional domains are refined based on the spatial correlation characteristics."

Then, the description of Fig. 1 and the second paragraph in Sect. 3.1. now reads as "As can be seen in Fig 1a, all of continental Europe seems to be well-correlated in terms of tropospheric ozone column monthly anomalies, while a further subdivision of East Asia, one of the regions defined in Wang et al. (2022), their Fig. 1, is required to define well-correlated regions (see Fig. 1b). In the latter case, the East Asia region shown in Fig. 1a was divided into three different well-correlated regions (e.g. regions 9, 10, and 11 in Fig. 3). In general, this correlation analysis method resulted in the refinement (i.e. splitting up) of many of the 11 original regions defined by Gaudel et al. (2020) and Wang et al. (2022): Europe (in Continental Europe, European Arctic), Western North America (Western USA, Pacific Northwest, California), East Asia (East China, Northeast Asia, South Japan), Southeast Asia (Southeast Asia, South China Sea), Malaysia/Indonesia (Indonesia, Southern Malay Peninsula), Persian Gulf (Persian Gulf, East Mediterranean Sea), Gulf of Guinea (Gulf of Guinea, West African Highlands), Northern South America (Middle America, Caribbean). The remaining three regions (Eastern North America, Southeast US, India) in Gaudel et al (2020) and Wang et al. (2022) did not need to be divided because all of their sites were well-correlated in terms of TrOC/FTOC. By adding some sites with different co-located techniques (Mauna Loa and Lauder) and the Canadian Arctic to the original regions defined by Gaudel et al. (2020) and Wang et al. (2022), we end up with 24 well-correlated regions that are covered by ground-based HEGIFTOM data. Those regions form the starting point for the regional trend estimations, for both TOST and the synthetized trend approaches, but the spatial and temporal sampling of the ground-based observations in those regions place further constraints on the final determination and selection of the regions."

And in Figure 1, I suggest showing the locations of IAGOS FRA airport and the FTIR Hefei station.

Done. Thank you very much for this good suggestion.

3) Page 10, Figure 3 –

Figure 3 shows the locations of 19 different regions. Were these regions identified by the spatial correlation analysis (which said 24 different regions)? Please clarify.

Yes. The 24 different regions were computed, as explained above, based on the correlation coefficients between the CAMS TrOC/FTOC at the HEGIFTOM site locations. However, the ground-based measurements themselves contain gaps. Only measurement sites with time series having at least 30 months of data were included in the regions, and only regions consisting of at least 2 sites fulfilling this data coverage criterion. In the end, 5 of those 24 regions therefore had to be discarded from our analysis.

In the text, we first specified that “In practice, the LMM method will be applied on the L1 (all measurements) data of the sites within one of the 24 regions defined in Sect. 3.1”. Then, we further wrote that “Because we included only sites with observation time series covering at least 30 months in the well-correlated regions, and we consider only regions consisting of at least 2 sites fulfilling this criterion, 5 of those 24 regions had to be discarded from the LMM trend estimation. The sites retained in each preserved region are listed in Tables 1 and S1, and shown in Fig. 3.”

4) Page 14, Line 350-357 -

For the TOST-based trends, 12 regions were selected and analyzed, as shown in Figure 4. It was not clear why the 19 regions as defined in Figure 3 were not used to facilitate the comparisons between TOST-based and HEGIFTOM-based trends. Please clarify.

The definition of the TOST regions was also based on the 24 regions from the correlation analysis, but, in contrast with the sample of sites used for the correlation analysis, TOST only deals with ozonesonde sites, not with IAGOS, FTIR, Umkehr, and lidar sites. So, the TOST sample to be used for defining regions is more limited, e.g. discerning sites with different techniques co-located, and moreover, the density of the TOST measurements (or the number of independent samples, see Figs. 4c and 4d) further limits the number of regions for the TOST trend estimations.

The text at the end of Sect. 4 has been modified to “Because the number of independent samples is (too) low in about half of the 24 well-correlated regions determined by the correlation analysis (see Section 3.1), we could select only 12 regions (Fig. 4a and Table 3) for analyzing TOST-based TrOC and FTOC trends. Within each region, high spatial correlation (>0.7, in line with Weatherhead et al., 2017) is found among stations. Table 3 lists the column-averaged mean independent samples over 1995-2021 in those 12 regions. The number and spatial extent of some of those regions is different from the (19) regions used for calculating LMM synthesized trends (see Sect. 3.2, Fig. 3 and Table 2), because these latter regions include sites or airports where ozone is measured with techniques or platforms other than ozonesondes (i.e. FTIR, IAGOS, Umkehr, and lidar).”

5) Figures 10-13

Figures 10-13 appear to need large revisions. The font size was difficult to read, and the font type was different from that of other figures, such as Figure 9.

Thank you. We revised Figures 10-13, increasing the font sizes, symbols, and thickness of the lines in the graphs. All figures in the manuscript now also use the same font type, Helvetica.

6) Since several regression methods were utilized in the study, a Table or some paragraphs summarizing their applications and differences would be helpful to smooth the paper structure. Right now, it is not clear in the manuscript why Quantile Regression was used in one case, while Dynamical Linear Modelling was used in another case. Would these regression methods return different findings?

Very good idea to provide such a table. We generalized the idea and created a table summarizing the main differences between the TOST and LMM approach, not only in terms of the statistics used, but also included the measurements used, the time coverage, how the regions have been defined, the merging method, etc. We hope that this table might also clarify some of the issues that you raised in earlier points.

This table can be found in the beginning of Sect. 5 (Trends) and the following text has been added to introduce it:

“In this section, we will present and discuss the trend estimates for the different well-correlated regions, optimized based on sampling density for each of the (gridded) TOST and (individual) HEGIFTOM ground-based sites. The specific properties of these two datasets calls for the use of different statistical trend estimation methods for determining regional trends: the TOAR-II recommended (Chang et al., 2023b) Quantile Regression (QR) and DLM (for a subset only) are applied to the merged gridded TOST data within a region, whereas the LMM utilizes a linear regression model to synthetize (or merge) the trends from the individual site time series in a region. The main characteristics of both datasets and the different statistical techniques are summarized in Table 4, which serves as the framework for this section.

Table 4. Summary of the most important differences between the two approaches that are used to calculate regional trends from the TOST and HEGIFTOM datasets.

	TOST	**LMM**
Measurements	Ozonesondes only, trajectory mapped	IAGOS, homogenized ozonesondes, FTIR, Umkehr, Lidar (HEGIFTOM)
Time coverage	1990-2021, 1995-2021, 2000-2021	1990-2022, 1995-2022, 2000-2022
Regions	N = 12 (see Fig. 4a and Table 3), based on correlation analysis (3.1) and density of independent samples (Figs. 4c & d)	N = 19 (see Fig. 3 and Table 1), based on correlation analysis (3.1) and availability of ≥ 2 sites with ≥ 30 months of data.
Merging method	Merging of gridded trajectory-mapped data and ozonesonde **measurements** (if available) within a region	Synthetizing **trends** from individual sites within a region
Trend estimation tools	QR trend estimation from the merged, regional, annual mean time series. DLM (Sect. 3.3) is used on a subset of regions using monthly mean (L3) time series (Fig. 8), refining the decadal trend estimation by showing how the trend varies over time	Linear regression model for calculating synthetized trends from well-correlated individual time series (L1, all measurements), allowing an intercept and a slope to adjust the difference from each individual trend against the overall trends (Sect. 3.2)

We also specifically added to the DLM description in Sect. 3.3: “DLM is only applied to three TOST regional monthly mean time series, to show how the trends for these regions changed over time.” From the yearly DLM trends, the decadal trends over the considered time periods 1995-2021 and 2000-2021 might be derived. However, in Van Malderen et al. (2025), we already showed that the DLM and QR trend estimates for the individual site time series were very similar. This was already included in Sect. 3.3.

7) Page 20, Figure 8

The legends p300 and FT need to be defined in the figure caption. DLM trends were presented here and in the Section 5.1.2, right? I do not see anywhere else described the DLM results (not in Sect 5.3 as said on Line 378).

We changed the legends to TrOC and FTOC, the acronyms that have been used throughout the manuscript to respectively denote the tropospheric ozone column between the surface to 300 hPa and the free-tropospheric ozone column between 700 and 300 hPa. DLM results are indeed only shown in Figure 8 and described in Section 5.1.2., while the DLM method itself is (briefly) described in Sect. 3.3 (and not 5.3) and referenced there for further details (Ball et al., 2017, Alsing et al., 2019).

8) Page 30, Line 667

“The difference is, at least for some regions, driven by the positive trends from measurement techniques other than ozonesondes.” What did the statement mean by trends from measurement techniques? Please better explain it.

TOST trends are only based on ozonesonde data, while the LMM method calculates synthetized trends from ozonesonde, FTIR, lidar, IAGOS, and Umkehr time series. If some of the densely sampled techniques (FTIR, lidar, Umkehr) display a strong (positive) trend in a region, the LMM trend is expected to be impacted considerably w.r.t. the TOST trend for the same region, without including data from those techniques. In the discussion of Sect. 5.3, some examples are given for the Europe and North America (except California) regions.

We have changed this sentence to “The difference is, at least for some regions, driven by the positive trends from the measurement techniques other than ozonesondes (FTIR, Umkehr, IAGOS, lidar) that have been included in LMM, but not in TOST.”

9) “TO BE FURTHER COMPLETED” in the Acknowledgements shall be noted.

Thank you, we completed the acknowledgements!