Inter-comparison of tropospheric ozone column datasets from combined nadir and limb satellite observations

Carlo Arosio[1], Viktoria Sofieva[2], Andrea Orfanoz-Cheuquelaf[1], Alexei Rozanov[1], Klaus-Peter Heue[3], Diego Loyola[3], Edward Malina[4], Ryan M. Stauffer[5], David Tarasick[6], Roeland Van Malderen[7], Jerry R. Ziemke[5], and Mark Weber[1]

[1]Institute of Environmental Physics, Bremen University, Bremen, Germany
[2]Finnish Meteorological Institute, Helsinki, Finland
[3]Institut für Methodik der Fernerkundung am Deutschen Zentrum für Luft- und Raumfahrt (DLR)
[4]European Space Agency, ESRIN, Frascati, Italy
[5]NASA Goddard Space Flight Center, Greenbelt, MD, USA
[6]Environment and Climate Change Canada, Downsview, ON, CA
[7]Royal Meteorological Institute of Belgium, Brussels, Belgium

Correspondence: Carlo Arosio (carloarosio@iup.physik.uni-bremen.de)

Abstract. This manuscript presents an inter-comparison between existing tropospheric ozone column (TrOC) datasets obtained using combined limb and nadir observations, i.e. exploiting collocated stratospheric profile and total column information retrieved from limb and nadir satellite observations, respectively. In particular, seven datasets have been considered, covering the past two decades and consisting of monthly averaged time series with nearly global coverage. We perform a comparison in terms of climatology and seasonality, investigate the tropopause height used for the construction of each dataset and the related biases, and finally discuss long-term TrOC ~~drift~~ drifts and trends. The overall goal of the study is to assess the consistency between the datasets and explore possible strategies to reconcile the differences between them. Despite uncertainties associated with the limb-nadir residual methodology and large biases between the mean values of the considered datasets, we ~~show~~ identify an overall agreement of TrOC ~~morphology~~distribution patterns. The different tropopause height definitions used to construct the datasets did not show a relevant role in explaining the biases between them. We demonstrate that a thorough investigation of the drifts with respect to ground-based observations is needed to evaluate TrOC trends from satellite data and that long-term trends in specific regions can be consistently detected, ~~as the~~ e.g. a positive trend of up to 1.5 DU per decade over ~~Southeast Asia~~China for the 2005-2021 period.

1 Introduction

Tropospheric ozone plays a crucial role in air quality and climate regulation, being a pollutant and a greenhouse gas (Warneck, 1999). It is responsible for respiratory problems, as it reduces lung function, and it has negative effects on ecosystems, e.g. inhibiting plant growth and reducing agricultural yields. Monitoring the concentration of tropospheric ozone is essential for understanding its impacts on health and climate ~~(Monks et al., 2015)~~(Mills et al., 2018; Fleming et al., 2018). Tropospheric ozone is a secondary pollutant ~~,~~produced by chemical reactions between nitrogen oxides (~~NOx~~NO_x) and volatile organic

compounds ($VOCs$) in the presence of sunlight ($\lambda < 420$ nm). The primary sources of tropospheric ozone precursors include vehicle emissions, industrial activities, and chemical solvents, making ozone a critical focus for air quality regulations and environmental health initiatives (Brown et al., 2013).

The direct retrieval of tropospheric ozone from nadir satellite observations has been investigated, starting with Global Ozone Monitoring Experiment (GOME) observations (Munro et al., 1998). More recently, tropospheric ozone retrievals have been performed, for example, using nadir measurements from GOME-2 (Miles et al., 2015), the Ozone Monitoring Instrument (OMI) (Bak et al., 2013), the Infrared Atmospheric Sounding Interferometer (IASI) (Boynard et al., 2009) and the TROPOspheric Monitoring Instrument (TROPOMI) (Mettig et al., 2021; Keppens et al., 2024). However, distinguishing between stratospheric and tropospheric ozone is challenging due to the low amount of ozone in the troposphere in comparison to the stratosphere, and the coarse vertical resolution of nadir observations.

Several residual methods involving the subtraction of the stratospheric ozone column (SOC) from the total column ozone (TCO) have been developed over the last decades. Cloud slicing is a technique that uses total column measurements taken at scenes with different cloud altitudes to derive vertical ozone profiles (Ziemke et al., 2001). The convective cloud differential (CCD) method is another technique used to measure tropospheric ozone (Ziemke et al., 1998). It exploits the presence of high convective clouds in the Pacific sector and assumes zonal invariance in stratospheric ozone. This method is particularly effective in tropical regions, where deep convective activity is frequent.

The limb-nadir tropospheric ozone residual (TOR) technique involves data from both limb-viewing and nadir-viewing satellite instruments, and provides tropospheric ozone column (TrOC) information. Limb-viewing instruments capture high-resolution vertical profiles by observing the atmosphere's limb, while nadir-viewing geometry offers higher horizontal sampling. This approach was first proposed by Fishman and Larsen (1987) for a combination of observations from the Total Ozone Mapping Spectrometer (TOMS) and the Stratospheric Aerosol and Gas Experiment (SAGE). In the last three decades this has been applied to several instrument combinations and this paper focuses on datasets retrieved using combinations of limb and nadir satellite observations (or reanalysis data) covering the last two decades.

Satellite observations over the past few decades have shown significant trends in tropospheric ozone levels (Gaudel et al., 2018). Regional variations are driven by factors such as anthropogenic emissions, including nitrogen oxides (NO_x) from industrial activity, and natural sources like wildfires and biogenic emissions. In many regions, especially in developing countries, tropospheric ozone levels have been increasing due to rising pollution levels, while in parts of North America and Europe ground ozone levels have stabilized or decreased thanks to air quality regulations (Pope et al., 2023, 2024). Understanding these trends is relevant for developing strategies to mitigate ozone pollution and to evaluate the success of such mitigation strategies, which may be costly. Despite the number of techniques, measurements and research conducted, reconciling the differences between satellite and in situ observations has been challenging (Tarasick et al., 2019).

This work fits within the framework of the Tropospheric Ozone Assessment Report (TOAR) II initiative, which has the aim of providing an up-to-date scientific assessment of the global distribution of tropospheric ozone and its trends from ground-based instruments and satellite observations. An overview of tropospheric ozone trends was given by Gaudel et al. (2018), concluding that satellite data were suitable for computing trends, but discrepancies between different datasets were large, with the need for further studies to reconcile those differences. Lately, a few studies tried to reconcile these discrepancies, such as Gaudel et al. (2024) for the tropics, pointing out the need to maintain and develop high-frequency continuous observations, and Pope et al. (2024) that show small linear trends at mid-latitudes affected by large inter-annual variability.

The manuscript is structured as follows: Sect. 2 describes the datasets used for this analysis and gives ~~some~~ more insights into the limb-nadir TOR technique. Sect. 3 presents a comparison of the selected TOR datasets ~~,~~ in terms of climatology, anomalies and seasonality. The role of the tropopause height definition used to construct the datasets is discussed in Sect. 4, where a method to correct the tropopause definition-related bias between time series is presented and assessed. In Sect. 5, a comparison with ozonesondes is presented; these independent measurements are also used to evaluate possible drifts in the datasets. Finally, in Sect. 6 the focus is on long-term trend studies, by using the datasets covering a period of at least 10 years~~of data~~.

2 Limb-nadir combined datasets

For this inter-comparison study we consider seven limb-nadir TOR datasets, as listed in Table 1, where the time frame of each product, the chosen tropopause height (TPH) definition and the horizontal resolution are listed. The choice of the TPH plays a relevant role for the construction of these datasets, as it is used as a bottom boundary for the integration of the stratospheric profile. If a gap between the two is present, a climatology or a model needs to be used to extend the profiles down to the TPH. Discrepancies in TPH are not linear in ozone ~~,~~ due to its increasing concentration with altitude in the stratosphere; in addition, the chosen TPH definition impacts ~~on~~ the sensitivity of the TrOC product to stratospheric ozone, as ~~its~~ the ozone profile generally starts to increase below the typical thermal tropopause (e.g. Monsees et al., 2024). Satellite datasets are in this study monthly Level 3 (L3, gridded) time series.

Table 1. Limb-nadir TOR datasets included in this study.

Name	Time frame	TPH definition	Original horizontal resolution
OMI-LIMB (Sofieva et al., 2022)	~~2004-2022~~ 2004-2023	Thermal (WMO) or ozonopause	1°x1°
GTO-LIMB (Sofieva et al., 2022)	~~2002-2022~~ 2004-2023	Thermal or ozonopause	1°x1°
S5P-BASCOE (Heue et al., 2022)	2018-2023	380 K (tropics), 3.5 PVU (extra-tropics)	0.25°x0.25°
SCIA+OMPS (Orfanoz-Cheuquelaf et al., 2023)	2004-2023	Thermal (tropics), 3.5 PVU (extra-tropics)	5°x5°
OMI-MLS (Ziemke et al., 2006)	2004-2023	Thermal	5°x5°
OMPS-MERRA (Ziemke et al., 2022)	2012-2023	2.5 PVU or 380 K (whichever lower)	1°x1°
EPIC-MERRA (Ziemke et al., 2022)	2015-2023	2.5 PVU or 380 K (whichever lower)	1°x1°

In the following the various datasets are briefly described.

OMI-LIMB and GTO-LIMB are two datasets developed at the Finnish Meteorological Institute (FMI) as part of the SUNLIT project (Sofieva et al., 2022) and within the ESA Climate Change Initiative (CCI). These datasets combine total ozone columns (TOC) either from NASA's OMI or from GOME-type Ozone (GTO) data (Coldewey-Egbers et al., 2022) with limb information coming from several instruments, such as MLS (Microwave Limb Sounder), OSIRIS (Optical Spectrograph and InfraRed Imaging System), MIPAS (Michelson Interferometer for Passive Atmospheric Sounding), SCIAMACHY (SCanning Imaging Spectrometer for Atmospheric CHartographY), OMPS-LP (Ozone Mapping and Profiles Suite – Limb Profiler), and GOMOS (Global Ozone Monitoring by Occultation of Stars). The ozone profiles were merged into a high-resolution dataset (LIMB-HIRES), which was used for computation of the stratospheric ozone columns (Sofieva et al., 2022). The MLS record is used as a reference in creating the LIMB-HIRES dataset. Two tropospheric ozone columns are provided in both OMI-LIMB and GTO-LIMB datasets: from ground to the thermal tropopause and from ground to 3 km below the thermal tropopause.

The S5P-BASCOE dataset (Heue et al., 2022) is generated through the combination of TROPOMI/Sentinel-5 Precursor (S5P) TOC (Garane et al., 2019) and the Belgian Assimilation System for Chemical ObsErvations (BASCOE), which assimilates ~~TROPOMI~~ MLS and other stratospheric ozone profiles into a chemical transport model to separate the stratospheric and tropospheric ozone components. The time series has a high spatial resolution and covers the period from 2018 to the present. An isentropic tropopause definition was used in the tropics, whereas in the extratropics a dynamical tropopause ~~is~~ was used.

The SCIA+OMPS dataset is the only merged product of the list, as the TrOC datasets independently retrieved for SCIA-MACHY and OMPS measurements have been merged on a monthly gridded basis. The retrieval of OMPS (Orfanoz-Cheuquelaf et al., 2023) and SCIAMACHY (Ebojie et al., 2016) TrOC data employs a specific TOR methodology called limb-nadir matching (LNM), which consists in combining two observations of nearly the same air mass performed by the same instrument (SCIAMACHY) or from two instruments on the same platform (OMPS). This technique has the advantage ~~to minimize~~ of minimizing instrument-related biases. The merging of the recently re-processed SCIAMACHY with LNM OMPS TrOC was performed on de-seasonalized anomalies of the two time series ~~,~~ by using OMI-MLS as a transfer function: the bias between SCIAMACHY and OMI-MLS anomalies in the period 2007-2012 and between OMPS and OMI-MLS over 2012-2017 was removed to merge them. OMPS TrOC seasonal cycle was then added back to get a ~~dataset covering 2002-2023~~ dataset in DU covering 2004-2023.

The OMI-MLS TrOC dataset (Ziemke et al., 2006) provides global measurements of tropospheric ozone by combining data from OMI and MLS, both aboard the Aura satellite. The dataset provides monthly L3 data from 60°S to 60°N at a resolution of 5° latitude by 5° longitude. To take into account an identified drift in OMI time series ~~(Gaudel et al., 2024), the time series~~ , this dataset was corrected by the data provider by adding a drift at a post-processing step, as described in Gaudel et al. (2024). This data spans from October 2004 to the present and was used in several studies to investigate long-term trends, pollution events, and interactions with other atmospheric gases (Ziemke et al., 2019, 2022). ~~TPH values are defined according to~~ The dataset adopts the WMO thermal ~~definition from~~ TPH definition (WMO, 1957), i.e. based on the 2 $\mathrm{K\,km^{-1}}$ thermal vertical gradient thershold, using NCEP reanalyses.

110   For the two OMPS-MERRA and EPIC-MERRA datasets, SOC information is derived by vertically integrating Global Modeling and Assimilation Office (GMAO) Modern-Era Retrospective analysis for Research and Applications-2 (MERRA-2) assimilated Aura MLS ozone profiles (Gelaro et al., 2017; Wargan et al., 2017), from the top of the atmosphere down to the tropopause. This is combined with TCO from OMPS and the Deep Space Climate ObserVatoRy (DSCOVR) EPIC (Marshak et al., 2018) instruments, respectively. The dynamical TPH ~~definitions~~ definition is used in both cases, with a 2.5 PVU or
115   380 K threshold, from MERRA-2 data.

All datasets are monthly averaged TrOC values and have been binned to the same spatial resolution, 5° latitude x 5° longitude, ~~and used in this resolution~~ for the following analysis.

3   Comparison of the datasets and their climatologies

A first comparison between the datasets is performed in terms of zonal averages to ~~assess their overall~~ get a rough assessment
120   of their biases, as shown in Fig. 1 for several latitude bands defined according to the TOAR II Guidelines. The figure showcases the presence of large discrepancies between the datasets, also in terms of seasonal cycle. For example, we notice the generally lower TrOC values from GTO-LIMB and OMI-LIMB, including ~~the~~ their more pronounced seasonal cycle at mid-latitudes with respect to the other datasets. Evident is also the different seasonal cycle shown by ~~the~~ OMI-MLS at southern mid-latitudes or the generally high bias of S5P-BASCOE TrOC above 40° latitude. The best absolute agreement between all datasets is
125   found during summer at northern mid-latitudes. Large discrepancies are visible at southern mid-latitudes, especially in the last three years of the time series.

These biases have several possible reasons: e.g., overall discrepancies in SOC and TOC between satellite products, the specific tropopause definition adopted to construct the TOR product, the climatology or model used to fill SOC gaps, and the criteria used for the combination of SOC and TOC. ~~In the Supplements, Fig. S1 provides a comparison of the annual mean~~
130   ~~climatologies. The overall biases between the datasets are discussed there as well.~~ For the rest of the paper, we aim to change the perspective to highlight similarities and possibly resolve the differences.

We then performed a comparison in terms of climatologies~~, as displayed in~~ . In the Supplements, Fig. S1 provides a comparison of the annual mean climatologies. Here Fig. 2 ~~,~~ displays mean climatologies of the datasets for each season. Values are averaged over each respective time period, so that discrepancies may also arise for this reason. To better assess
135   the common ozone patterns, the climatologies have been de-biased ~~,~~ with respect to the multi-instrumental mean so that the global mean value [60°S, 60°N] for all maps in each column is the same (and reported on top). We focus here on common features of TrOC that are evident for all datasets, in particular, the wave-1 pattern in the tropics, as visible in Fig. 2. This is a clear zonal asymmetry in the TrOC distribution with higher ozone concentrations over the Atlantic and African regions, and lower concentrations over the Pacific and Indian Oceans (Fishman et al., 1990). The main reasons for this pattern are related
140   to the large biomass burning taking place in the African and South American regions, but also the weaker intensity of deep convection in the Atlantic sector. This pattern is influenced by ~~the~~ El Niño–Southern Oscillation (ENSO) and by the intensity of the Intertropical Convergence Zone (ITCZ) (Bruckner et al., 2024). Regional hot-spots of TrOC are visible over polluted areas,
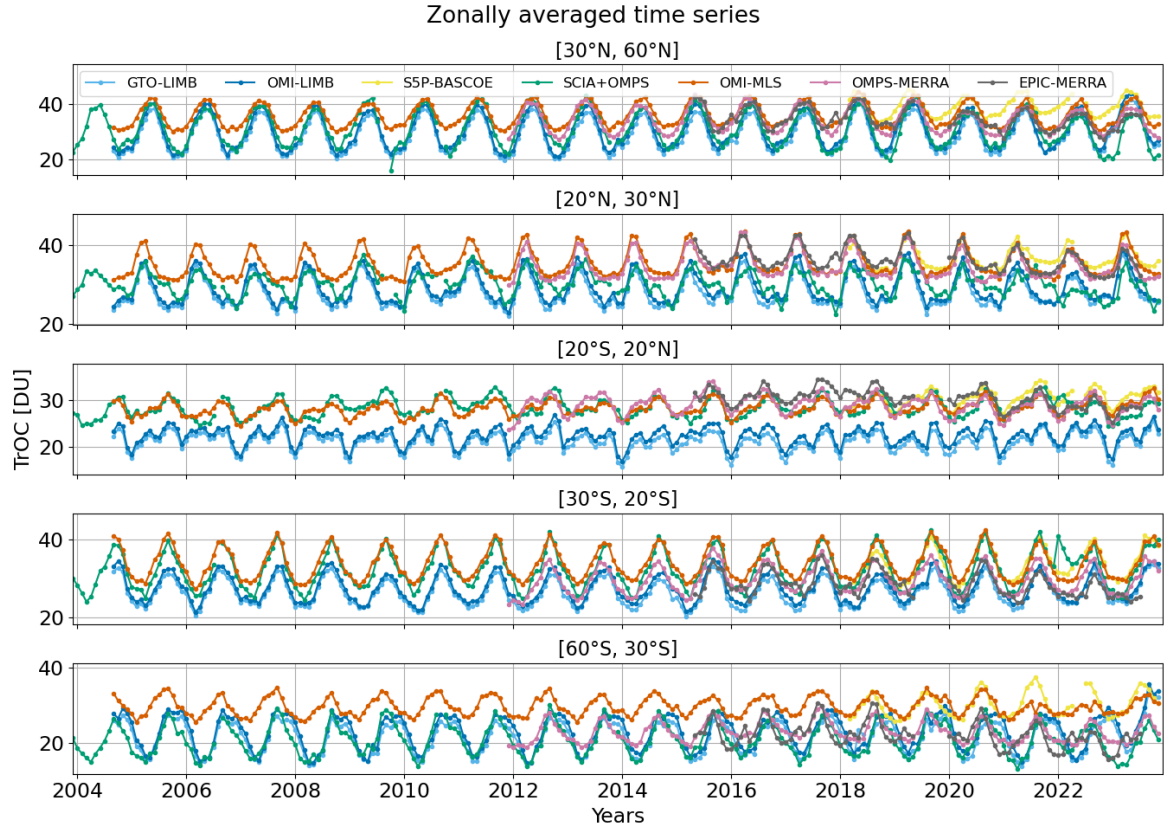
Figure 1. Time series of zonal mean TrOC averaged in several latitude bands.

where precursor emissions are high, such as over China, Southern Europe, between the Arabian sea and India, and the west coast of US. Low ozone concentrations are typical over the oceans and unpolluted regions due to limited precursor availability.

145 Comparing the rows of the figure, we notice datasets displaying a large ozone seasonality, e.g. OMI-LIMB and GTO-LIMB that show the largest summer TrOC values at northern mid-latitudes. S5P-BASCOE shows higher ozone values in winter at northern ~~high-latitudes~~high latitudes, whereas EPIC-MERRA has unusually low ones at southern ~~high-latitudes~~high latitudes. In the tropics, SCIA+OMPS and OMPS-MERRA display less pronounced seasonality and longitudinal structure with respect, for example, to OMI-MLS or OMI-LIMB.

150 The seasonality is further investigated in Fig. 3 by plotting the seasonal cycle of the datasets averaged in several latitude bands. The mean seasonal cycle is defined as follows:

$$SC_m = \frac{1}{N_m} \sum \nolimits_{j=1}^{N_m} TrOC(t_j) \sum\nolimits_{t=1}^{N_m} TrOC(t), \qquad (1)$$
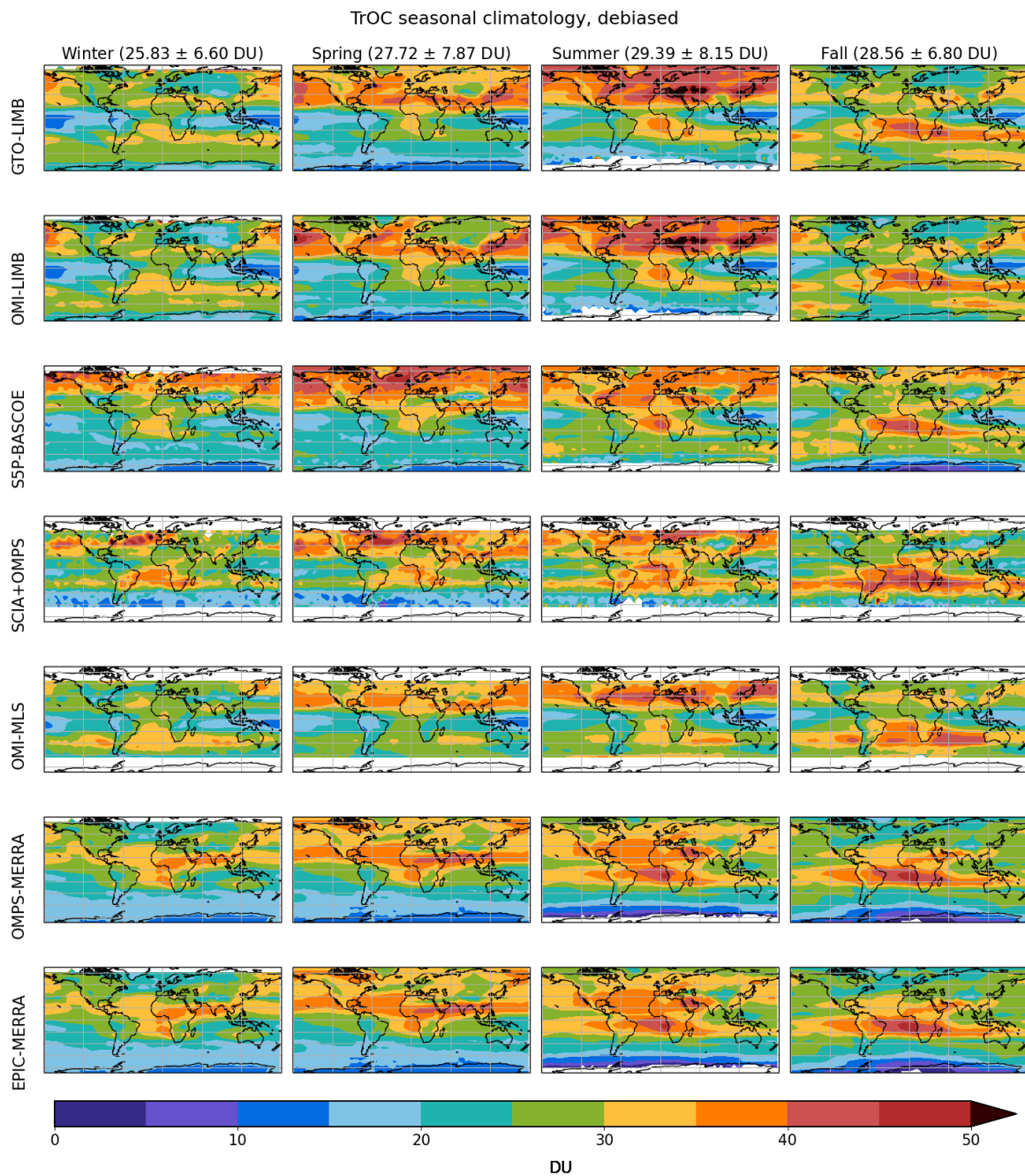
Figure 2. Climatology of the considered datasets, averaged over each season and after debiasing them to the same mean value, i.e. after bringing them to the ~~same~~ multi-instrument global mean. Titles include multi-instrument mean TrOC values and corresponding standard deviations for each season.

where N_m is the number of available monthly mean values ~~$TrOC(t_j)$ for the~~ $TrOC(t)$ for each specific month of the year m, e.g. January, in each time series. The offsets between the datasets have been removed, bringing all seasonal cycles to the same yearly average value, corresponding to the satellite-ensemble average, for a better comparison.
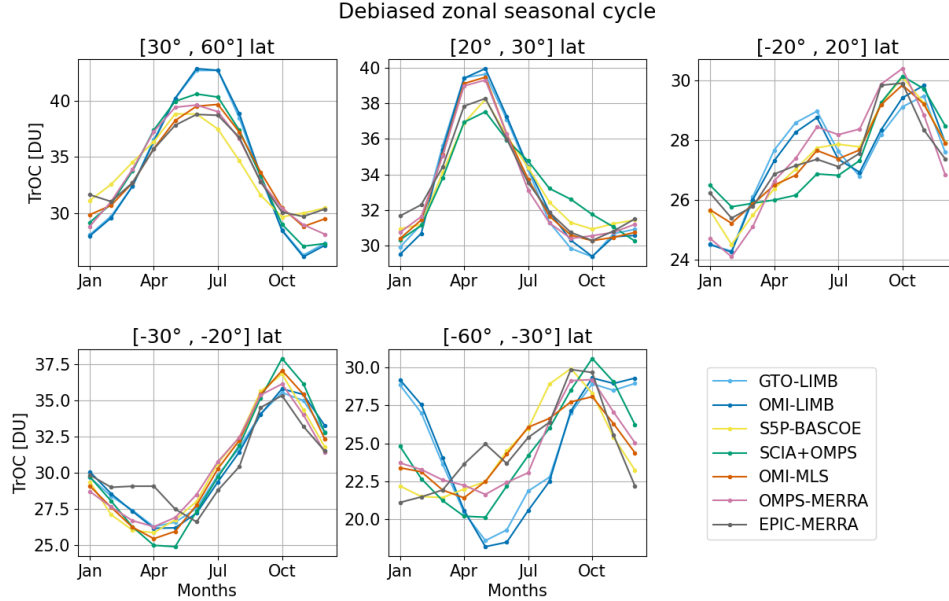
Figure 3. Seasonal cycle of the seven considered datasets (over their respective time coverage) in several latitude bands, after de-biasing to the multi-instrument mean for each panel.

~~Tropospheric ozone levels exhibit strong seasonal cycles and vary by location.~~ TrOC values are generally higher in the Northern Hemisphere during summer due to increased photochemical production and lower in winter. In the Southern Hemisphere, seasonal patterns are less pronounced. We notice that at most latitude bins the zonally averaged seasonality is in very good agreement between the datasets. The worst agreement is visible at southern mid-latitudes, with OMI-LIMB and GTO-LIMB showing a stronger seasonal cycle with respect to the others. In addition, ~~an overall~~ a large scatter is visible in ~~spring-time~~austral winter-time (JJA).

Another comparison related to the temporal evolution of the data products is shown in Fig. 4, which displays the time series of the absolute anomalies as a function of latitude, which are defined as follows:

$$\Delta(t_m) = TrOC(t_m) - SC_m, \tag{2}$$

where m indicates the month of the year, e.g. January, and t_m all months~~m~~, e.g. all Januaries, in the time series.

From Fig. 4, we can assess the presence of outliers in the data, as well as discontinuities or anomalous periods. In this respect, a pronounced feature is related to the drop that occurred in 2020, with negative anomalies up to 3-4 DU visible, especially in OMPS-MERRA and EPIC-MERRA, but also present in the other datasets with a smaller magnitude (1-2 DU). Ziemke
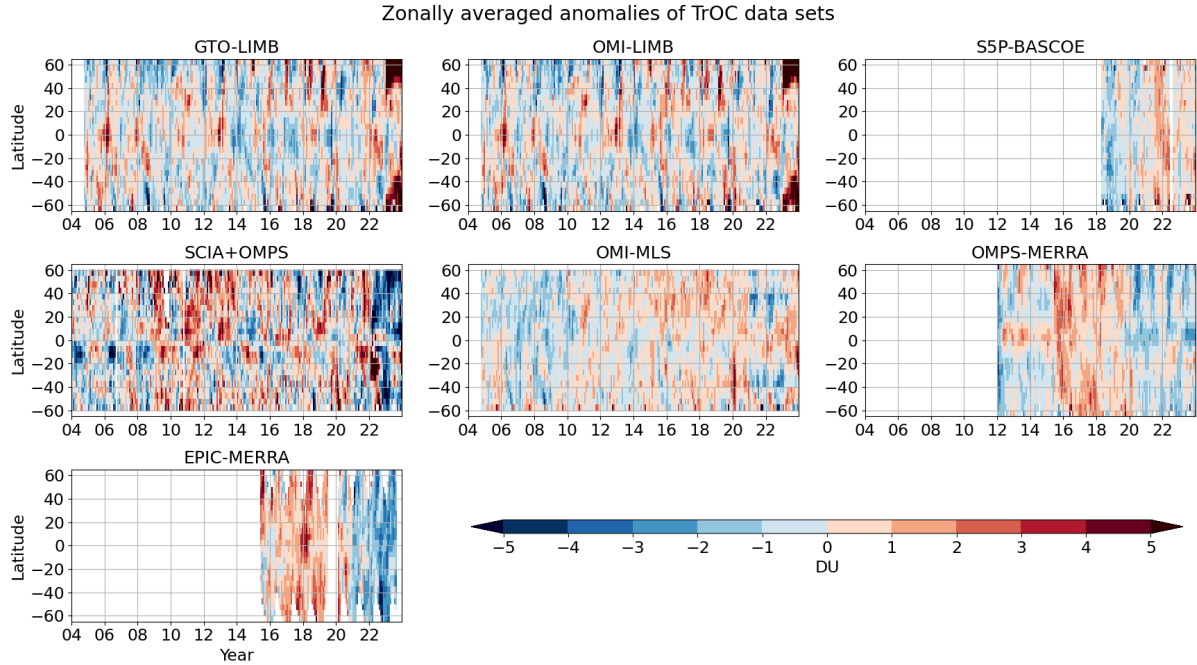
Figure 4. De-seasonalized time series (with respect to each respective period) as a function of latitude for the seven considered datasets.

et al. (2022) described a drop in ozone values at northern mid-latitudes starting from 2020 and attributed it to the COVID-19 pandemic. Ground-based observations have detected a small drop at some stations of up to 1-2 DU (Steinbrecht et al., 2021), with hints of a rebound over North America after 2021 (Chang et al., 2023a). ~~The presence of a such larger drop in some datasets is still under investigation. We looked for the presence of any sudden change in reanalysis temperature and stratospheric ozone without finding convincing evidences of this pattern (See Fig. S2 in the Supplements). We also~~ Other studies recently addressed this topic, providing the chemical background to understand this drop. Miyazaki et al. (2021) estimated an ozone loss of up to 5 ppb in the lock-down months in 2020 due to the reduction in NO_x anthropogenic emission. Putero et al. (2023) found negative ozone anomalies at high-elevation sites in North America and western Europe, particularly in 2020, confirmed also by Chang et al. (2022), Elshorbany et al. (2024) and Pimlott et al. (2024).

In our datasets, we found that when changing the period for calculating the seasonal cycle, this drop becomes evident in other datasets as well: Fig. ~~S3~~ S2 shows anomalies as in Fig. 4, but with the seasonal cycle computed over 2016-2019, instead of using the entire period of the respective time series. Fig. ~~S4~~ S3 better showcases the negative anomalies (example time series at 42.5°N) in recent winters also for GTO-LIMB ~~,~~ and OMI-LIMB ~~, and SCIA+OMPS~~ when the 2016-2019 is subtracted. ~~Finally, Fig. S5 displays the TPH anomalies:~~ The presence of a such large drop in some of the considered datasets is still under investigation. We looked for the presence of any sudden change in reanalysis temperature and stratospheric ozone without finding convincing evidences of artefacts that could explain this pattern. Also in TPH anomalies no evident discontinuities

have been found around ~~2020.~~ 2020 (as shown Fig. S4), ruling out a possible TPH-related artifact. Further studies are needed to better understand this feature and the discrepancies between datasets.

Apart from this drop, a high-ozone artefact related to the Hunga-Tonga volcanic eruption is visible in Fig. 4, particularly in SCIA+OMPS data, most probably coming from a sub-optimal filtering of the stratospheric column data. This dataset also shows a noisier time series than the others. A Quasi Biennial Oscillation signature ~~is evident,~~ was detected particularly in GTO-LIMB and OMI-LIMB datasets, which is most probably a SOC ~~signal~~interference. This signature is not well visible from this figure but is shortly discussed in Sect. 6. A positive TrOC trend can be identified by eye in the OMI-MLS dataset, but this is not evident for other datasets. Unexpectedly large anomalies in SCIA+OMPS, GTO-LIMB and OMI-LIMB in 2023 are under investigation.

4   Role of the TPH discrepancies and possible corrections

The definition of TPH used in the construction of the datasets, listed in Table 1, ~~plays~~ may play an important role in the biases between them. As already mentioned, discrepancies in TPH influence TrOC ~~,~~ due to the strong ozone gradient at this altitude. Figure 5 shows the zonal averaged values of the TPH (in hPa) for the selected datasets, each over the respective time frame. ~~Difference~~ Differences are largest in the sub-tropical and mid-latitude regions, particularly for OMPS-MERRA (and EPIC-MERRA, using the same TPH data), which adopts a dynamical definition of 3.5 PVU.

~~The difference between the datasets using the thermal definition tropopause, e.g. SCIA+OMPS and OMI-LIMB in the tropics, is related to the different definition of the thermal tropopause, according to Reichler et al. (2003) or WMO (1957).~~ For some datasets, such as SCIA+OMPS, another impacting factor is related to the climatology used to complete the stratospheric limb profiles in case the lowest measurement point lies above the TPH. In this way, depending on the limb vertical range and the used TPH definition, the adopted climatology will play a more or less relevant role.

It is clear from comparing this figure with Fig. 1 that the main biases between TrOC are not related to the tropopause definition. For example, OMPS-MERRA and EPIC-MERRA have the lowest ~~tropopause height~~ TPH at northern mid-latitudes, whereas the tropospheric ozone column is the largest, opposite to expectation. ~~However~~Nevertheless, we investigated an approach to remove the biases between the datasets related to the different TPH definitions by subtracting the ozone sub-columns corresponding to the gap in TPH. This was done using the European Centre for Medium-Range Weather Forecasts (ECMWF) ERA5 monthly gridded ozone profiles at each latitude and longitude bin. As reference TPH, the ERA5 dataset was selected. The chosen time period for this analysis is 2018-2022 as it is covered by all datasets. Figure 6 shows in panels (a) and (c) an example of TrOC time series for the selected datasets over the chosen period, zonally averaged at 45°N, respectively before and after subtracting the column gaps displayed in panel (b). We subtract the ~~mean~~ column gaps averaged over the selected period, ~~which are displayed in Fig. S6, right panel~~for each latitude/longitude bin. As one can see in the bottom panels, the standard deviation of the mean values of the datasets in each latitude-longitude ~~bins~~ bin generally increases after applying the correction. This is caused, as seen comparing ~~panel~~ panels (a) and (c), by the ~~large~~ larger correction needed for the OMPS-MERRA and EPIC-MERRA lines, bringing them to a lower level with respect to the other time series. An exception is the
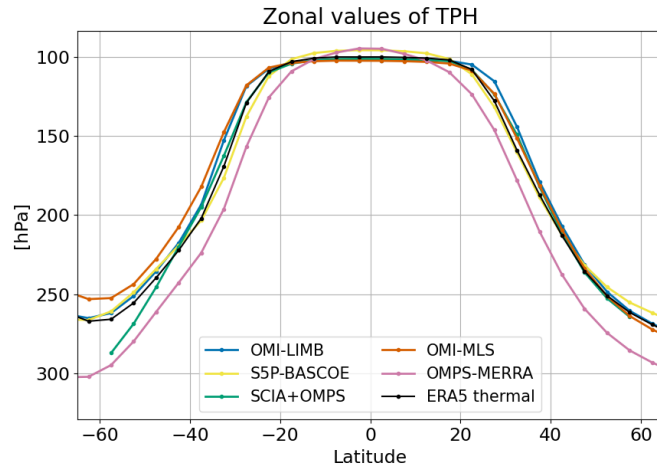
Figure 5. Zonal mean TPH (pressure, hPa) for the considered datasets, including ERA5 thermal tropopause values.

northern subtropical band, where the standard deviation after the correction decreases, indicating that, at least in this region, the biases between datasets are mostly caused by TPH discrepancies.

220   These results show that this methodology does not consistently reduce the bias between the time series. We also tested the removal of monthly-resolved TPH-related biases, i.e. removing seasonal cycle biases, ~~as shown in Fig. S7,~~ without improving the overall agreement of the datasets either. It would be useful to test this correction on L2 data, which is however out of the scope of this study, as we ~~analyize~~ analyze L3 data only. However, in Fig. ~~S8~~ S5, we show a hint of the better agreement that we could obtain with a L2 correction ~~,~~ using SCIAMACHY TrOC data.

225   Another important aspect of the ~~used~~ TPH having an impact on the TOR product is related to their possible long-term drift. To assess this aspect, we computed linear trends of the deseasonalized anomalies of the TPH time series, zonally averaged over the period 2005-2021. Figure 7 displays these linear trends for the products with the longer time series, as reported in the legend, including the ERA5 thermal tropopause data. Shaded areas are the 2-σ uncertainties ~~, i.e. corresponding to the 95 % confidence level~~. GTO-LIMB is not included as the results are very similar to those of OMI-LIMB. Trends are generally within 230 ± 3 hPa per decade, with ~~fairly common structures~~ a fairly common pattern as a function of latitude ~~, although~~ for the different datasets, except for the southern mid-latitudes, where a larger scatter is found and even the sign of the trend changes between the datasets. These differences are possibly related to different sampling of the satellite observations at these latitudes, to the different TPH definitions, and possibly to a discrepancy between ERA5 and MERRA-2 reanalysis. In addition, SCIA+OMPS is showing some larger deviations from the others due to sampling issues. ~~This indicates that at most locations, we can expect 235   a non-zero TPH trend, which should be taken into consideration when assessing TrOC trends from the datasets.~~

In the bottom panel of Fig. 7, the TrOC trends ~~only due to the TPH trends~~ (due to TPH trends only (from panel a) are determined using ozone data from ERA5 ~~. In the first step~~, following these three steps:
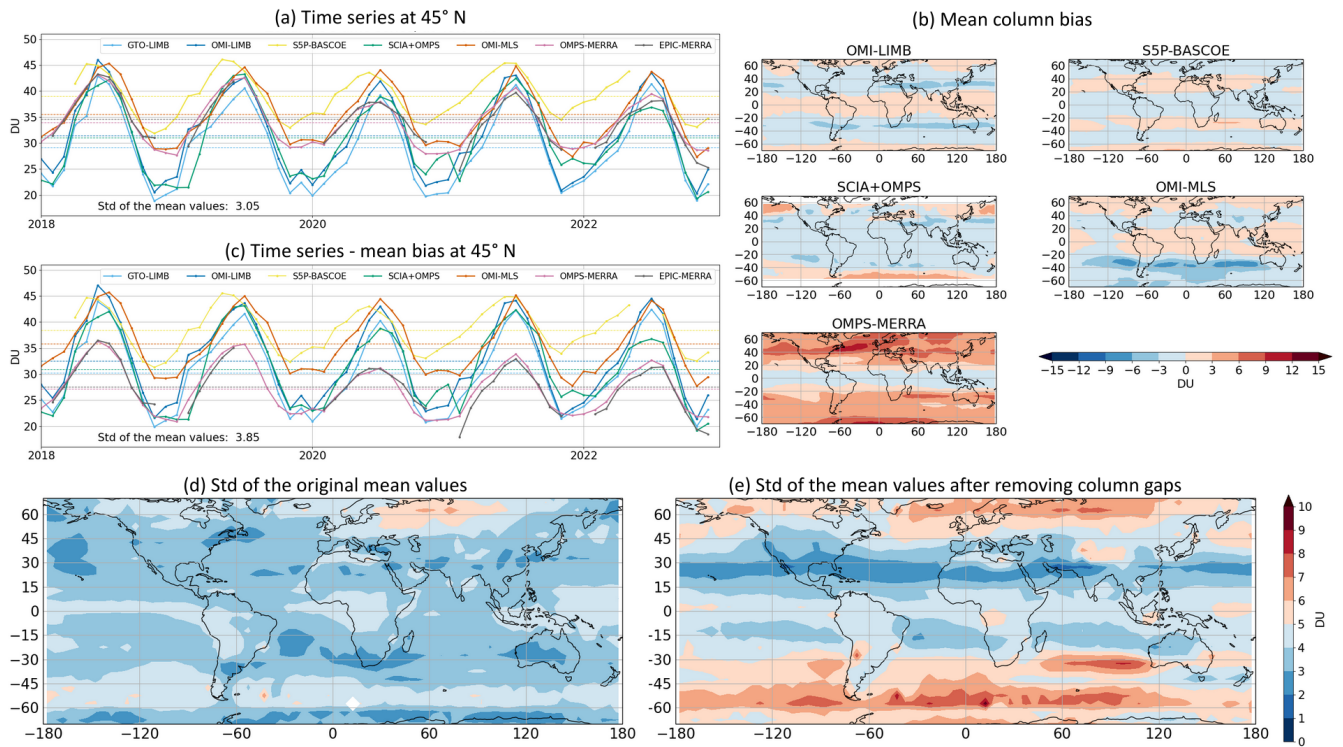
Figure 6. Panels (a) and (c) show an example of TrOC time series from 2018 to 2022, zonally averaged at 45°N, respectively before and after subtracting the column gaps displayed in panel (b). The bottom panels (d) and (e) show the standard deviation of the mean of the datasets in each bin, respectively before and after applying the correction.

1. The ERA5 TrOC time series ~~were derived using the thermal definition of the ERA5 TPH . In a second step,~~ was derived by integrating the ozone profiles up to the corresponding thermal TPH (from ERA5 ~~TrOC time series were re-calculated by using adjusted TPHs by adding the~~ temperature profiles).

2. TPH trends from the individual datasets (from panel a) were added to the ERA5 TPH ~~. The trend of~~ time series and the ERA5 TrOC time series was re-calculated by using the adjusted TPHs values.

3. The linear fit to the difference between both ERA5 TrOC time series was calculated and is shown in panel ~~b. The TPH related~~ (b).

The TPH-related TrOC trends are generally ~~non significant, except~~ close to zero in the tropics, within ±0.5 DU per decade in the [-30°S, 30°N] band. On the contrary, in the sub-tropics and at mid-latitudes ~~where they reach up to~~ both values and the discrepancy between the datasets get larger, with a peak around 37.5°N of about +1 DU per decade. These results should be taken into consideration when assessing TrOC trends from the datasets.
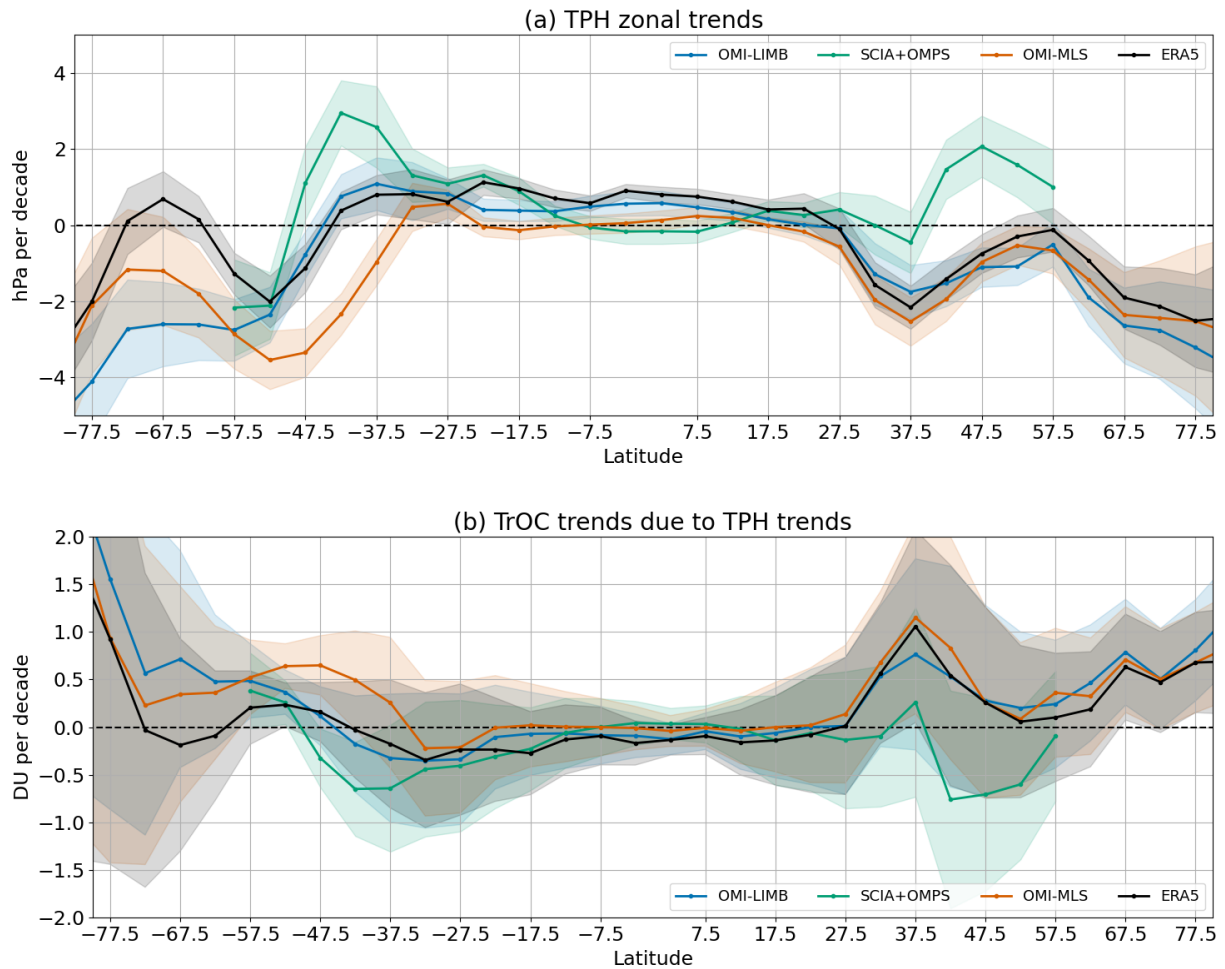
Figure 7. Linear trends in hPa per decade of the TPH from the datasets, including ERA5 monthly time series, as a function of latitude. In the bottom panel the respective TrOC trends are shown in DU per decade.

5 Comparison with HEGIFTOM sondes and relative drift

The HEGIFTOM (Harmonization and Evaluation of Ground-based Instruments for Free Tropospheric Ozone Measurements) working group ~~aim~~ aims at evaluating and harmonizing tropospheric ozone data obtained from different observing networks of ground-based instruments ~~,~~ in order to reconcile the differences in ozone distribution and trends between the different ground-based platforms (Van Malderen and Smit, 2020). For the present analysis, we are using monthly mean values of TrOC provided for the sonde stations listed in Table S1 in the Supplements~~, after the harmonization of their time series by the HEGIFTOM group~~. The TrOC was derived by integrating the sonde profiles from ground to the thermal tropopause level~~.~~

~~An overall inter-comparison between HEGIFTOM and satellite data to provide a general assessment of the absolute bias and scatter between them and their trends is described in Appendix A~~, which was derived by the data provider from the ozonesonde temperature profiles using the WMO definition.

For the present study~~however~~, the main aim of using the HEGIFTOM TrOC time series is to provide an assessment of potential drifts affecting the satellite TrOC datasets. ~~Since the averaging kernels (AK) of TOC retrievals typically show a lower sensitivity to the lowermost troposphere, we tested the role of the close-to-surface ozone in terms of overall biases with the satellite products and the effects on the drift analysis (and trends, see Fig. B2). The application of satellite-specific AK is beyond the scope of this work; however, to take into account the different vertical sensitivity of the TOC measurements with respect to high-resolved ozonesondes, we tested a simple weighting of the sonde profiles before vertical integration, as described by a function providing a rough approximation of OMI averaging kernels:~~

$$w = \left(\frac{z}{z_{lim}} \right)^3 \text{ for z} < \text{z}_{lim} \,,$$

~~where z is the sonde altitude, z_{lim} is one third of the TPH and the weight w is equal to 1 above this altitude. This weighting does not conserve the TrOC value, as we can see in Fig. S9, which shows comparison in absolute ozone values of the satellite datasets collocated at four sonde station: Hohenpeissenberg, Madrid, Hilo and Natal. The application of the weighting approach leads to an overall reduction of the high summer-time values~~The drift is defined as the linear trend of the difference between the satellite data and the reference HEGIFTOM time series.

For the drift assessment~~we used the weighted HEGIFTOM time series. The procedure is as follows: after applying the weighting described in Eq. A1,~~, we followed the procedure described here: for each available HEGITFTOM sonde station (see Table S1), we found the corresponding ~~collocated grid cell of each satellite dataset and we~~ satellite data grid cell containing the location of the station and computed de-seasonalized (absolute) anomalies for both time series. ~~Some stations were not considered due to the low number of available observations or for their short time coverage.~~ For each station, we then computed differences between anomalies (TOR - HEGIFTOM). The linear trend of these differences ~~is defined as~~ corresponds, as already mentioned, to the drift of the satellite product with respect to ~~the~~ sonde observations. To minimize the noise, we focused on two latitude bands, i.e. tropics [-30°S, 30°N] and northern mid-latitudes [40°N, 60°N], where enough sonde stations are available for taking the mean. We discarded sonde stations with a particularly short or sparse record ~~,~~ and satellite datasets having a short time span, i.e. S5P-BASCOE and EPIC-MERRA. For the ~~other~~ remaining TrOC products, Fig. 8 shows the time series of the differences and their respective linear trends, i.e. the drifts of the satellite datasets, averaged over the two selected latitude bands. Both panels display the monthly time series and their 13-month running averages for a less noisy visualization of the long-term tendency.

~~As we can see, the drifts~~At northern mid-latitudes (panel a) the drifts for three datasets, i.e. OMI-LIMB, GTO-LIMB and OMI-MLS are fairly close to zero~~but still significantly different (at 2σ) for most datasets.In particular, in the tropics~~, with p-values > 0.33, which corresponds to very low probability of a drift. SCIA+OMPS is affected by larger oscillations in the difference to HEGIFTOM and shows a negative drift, still with a p-value > 0.1, i.e. low confidence of a drift. OMPS-MERRA has only an 11-year time span, over which it displays "highly certain" drift.
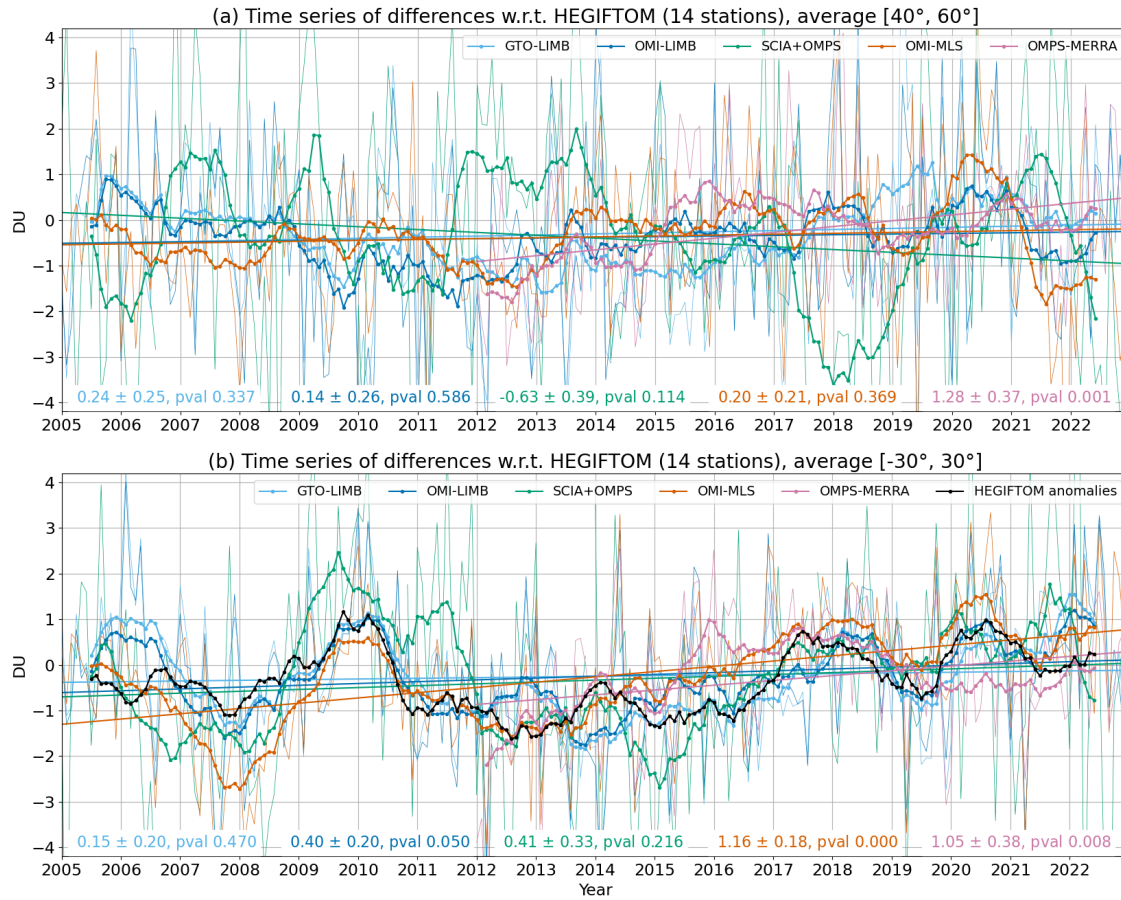
Figure 8. ~~Time series and linear trend~~ Drift in ~~DU per decade~~ DU/dec of the ~~differences~~ difference time series between satellite and HEGIFTOM sonde anomalies~~(after weighting)~~. Thin lines indicate the monthly difference time series, whereas thicker lines are their 13-month averages. ~~Trend values~~; in panel (~~drifts~~b) the HEGIFTOM anomalies are also shown (with sign changed). Drift values with corresponding 2σ uncertainties and p-values are reported at the bottom of the panels. The titles include the number of stations used for the analysis.

290   In the tropics (panel b) we notice a common ~~long-term behaviour, with negative trends until around 2014 and positive trends in the last 10 years.This might be related to the drop in ozone column from ozonesonde observations particularly in the tropics (e.g. Stauffer et al., 2022). This explains the large drift~~value derived for the OMPS-MERRA product, which covers only the last pattern in the differences among the five datasets, all affected by a positive drift with pronounced oscillations. For OMI-MLS and OMPS-MERRA the drift has a p-value < 0.01 (i.e. "very high confidence" of a drift), especially over the last 10 years.

295   ~~This artefact is reduced by the weighting introduced with Eq.A1: in the Supplements Fig. S10, the same pictures for the drift analysis without weighting display a very evident similarity between the sonde anomalies and the differences between satellite data and these anomalies, indicating that this pattern originates~~ OMI-LIMB shows the lowest drift with a p-value > 0.33, i.e.

"very low confidence". Over-plotted in black is the HEGIFTOM anomaly time series, changed of sign: the similarity between this line and the satellite-to-sonde differences indicates that these patterns are not captured by the satellite datasets. We further investigated the presence of these patterns by applying a weighting to the ozonesondes with the aim to test the influence of the lowermost troposphere on these patterns, as we know that the sensitivity of nadir observations decreases in the lowermost troposphere (Sofieva et al., 2022). This is discussed in Appendix A. Fig.A1 shows that the pattern described in the tropics is partially reduced by reducing the contribution from the lowermost troposphere. ~~At northern mid-latitudes, the~~ In this case, drifts are closer to zero (except for SCIA+OMPS~~dataset shows on average some larger deviations from HEGIFTOM in three time windows: 2009, 2012-13 and in 2018, which are under investigation. All other satellite datasets show a small positive drift~~) with p-values larger than 0.15 for OMPS-MERRA, OMI-LIMB and GTO-LIMB. This suggests that for investigating the stability of satellite-based time series with respect to ~~ozonesondes since 2005.~~ high vertical resolution observations, the application of averaging kernels could be beneficial, to take into account the different sensitivity with respect to satellite measurements.

An overall inter-comparison between HEGIFTOM and satellite data to provide a general assessment of the absolute bias and scatter between them and their trends is described in Appendix B.

6 Trends in geographical regions

~~Studies on free tropospheric~~ Studies on tropospheric ozone trends from ozonesonde measurements, such as Christiansen et al. (2022) ~~, Wang et al. (2022) , Stauffer et al. (2024), and ? highlighted the presence of small trends of +(1~~ analyzing trends over the ~~1990- 4)%/decade, depending on the selected station, with significant positive trends seen only in the Southeast Asia~~ 2017 period and Wang et al. (2022) over the 1995-2017 period found generally positive trends in the free troposphere and larger positive values in the lower troposphere in Southeast Asia, also confirmed by Stauffer et al. (2024) over the 1998–2022 period. Recently, Van Malderen et al. (2025) performed a thorough analysis of the HEGIFTOM dataset, assessing trends over the period 2000-2022 from ozonesonde stations globally distributed and comparing different trend calculation methods. The authors concluded that TrOC trends generally lie within the [-3 ppb/dec; +3 ppb/dec], with difficulties in finding common consistent geographical patterns.

Long-term satellite-detected ozone trends were discussed recently by several studies. Pope et al. (2023) investigated changes in the lower tropospheric column over the period 1996-2017, finding positive trends with high confidence in the tropics and with lower confidence at mid-latitudes. Froidevaux et al. (2025) showed that MLS-based upper tropospheric trends over the 2005-2020 period are consistent with TrOC trends in the tropics (Ziemke et al., 2019), with positive values over Southeast Asia and the sub-tropical Atlantic region.

We explored the changes in TrOC from three satellite data records, i.e. OMI-LIMB~~(or GTO-LIMB)~~, SCIA+OMPS and OMI-MLS, which cover the longest time frame. The chosen period is 2005-2021 to avoid possible perturbations from the Hunga-Tonga eruption in 2022 and unexplained features in some datasets in 2023 (as pointed out in Sect. 3). GTO-LIMB has shown very similar trends to OMI-LIMB and ~~is not reported~~ has not been included in the next figures. Instead of considering

latitudinal averages, we ~~considered~~ focused on specific geographical regions, which are of interest ~~for~~ to human-related activities and their changes over the last decades. In order to define the extent of these regions, we investigated the spatial correlation of the seasonal cycle of the chosen datasets, requiring a high homogeneity of the seasonality within each region and between the datasets. An example is shown in Fig. ~~S11~~S6.

Figure 9 displays the defined geographical regions of interest, which are:

- US, where stringent policies where introduced to reduce air pollution;

- the Mediterranean region, typically affected by high summer levels of ozone;

- ~~Southeast Asia~~China, where we expect an increase in TrOC values as reported by the sondes;

- the Atlantic Ocean off the African coast, a region affected by transport of precursors from wild fires and characterized by high levels of tropospheric ozone;

- the Amazon area, because of the change in land use and deforestation.
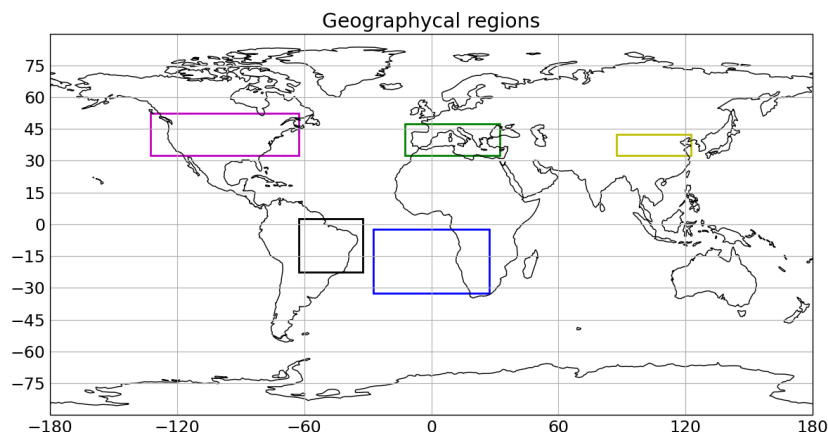
Figure 9. Geographical regions selected for the trend study.

~~To~~ From the time series averaged in each geographical region, we removed the seasonal cycle (Eq. 2, to get absolute anomalies) and we applied a ~~multivariate regression model, which includes~~ quantile regression (QR) model, as recommended by TOAR-II to reduce the impact of outliers on the regressed trends. In the regression model, we included three proxies: the first two principal components of the QBO and the Multivariate El Nino Southern Oscillation (ENSO) Index (MEI), without any lag. The latter in particular has a relevant impact on ozone variations in the tropical regions (e.g. Rowlinson et al., 2019). We include the QBO to account for its ~~signal contamination from the SOC. TOAR-II recommends the use of a quantile regression (QR) model to reduce the impact of outliers on the regressed trends. In this study no proxies were included in the QR. We compared the results of implementing the two regression models and concluded that the usage of~~ influence on the SOC: Fig. S7 shows the fit contributions of these two proxies in the ~~model is important, and that the difference between MLR and QR results are less~~

~~relevant for these monthly averaged datasets (not many outliers as in daily data). This is shown in Fig. ?? (Appendix B) and in Fig. S12, which compares trend time series from~~ tropics for OMI-LIMB data ~~using three regression models: MLR, QR, and simple linear fit. In addition, Fig. S13 provides an example of the role of the chosen proxies for the ozone time series fitting in specific regions.~~ We also tested the use of different proxies, such as the aerosol optical thickness and the TPH time series~~.~~

355 ~~Results are compared in Fig. S14 and do not show~~, without finding a significant contribution of these proxies to the trends.

Figure 10 shows the absolute anomaly time series and the respective trends in the five defined regions for OMI-LIMB, SCIA+OMPS and OMI-MLS, over the period ~~2004-2022~~2005-2021. The subplots show in thicker lines the 13-month running mean of the dataset time series averaged within each defined region ~~,~~ after subtracting the fit contributions from QBO and ENSO. The respective ~~MLR~~ QR trend values and uncertainties are also reported. Table 2 summarizes the results per region and reports the trend values in ppb per decade as well [1], with p-values and confidence.
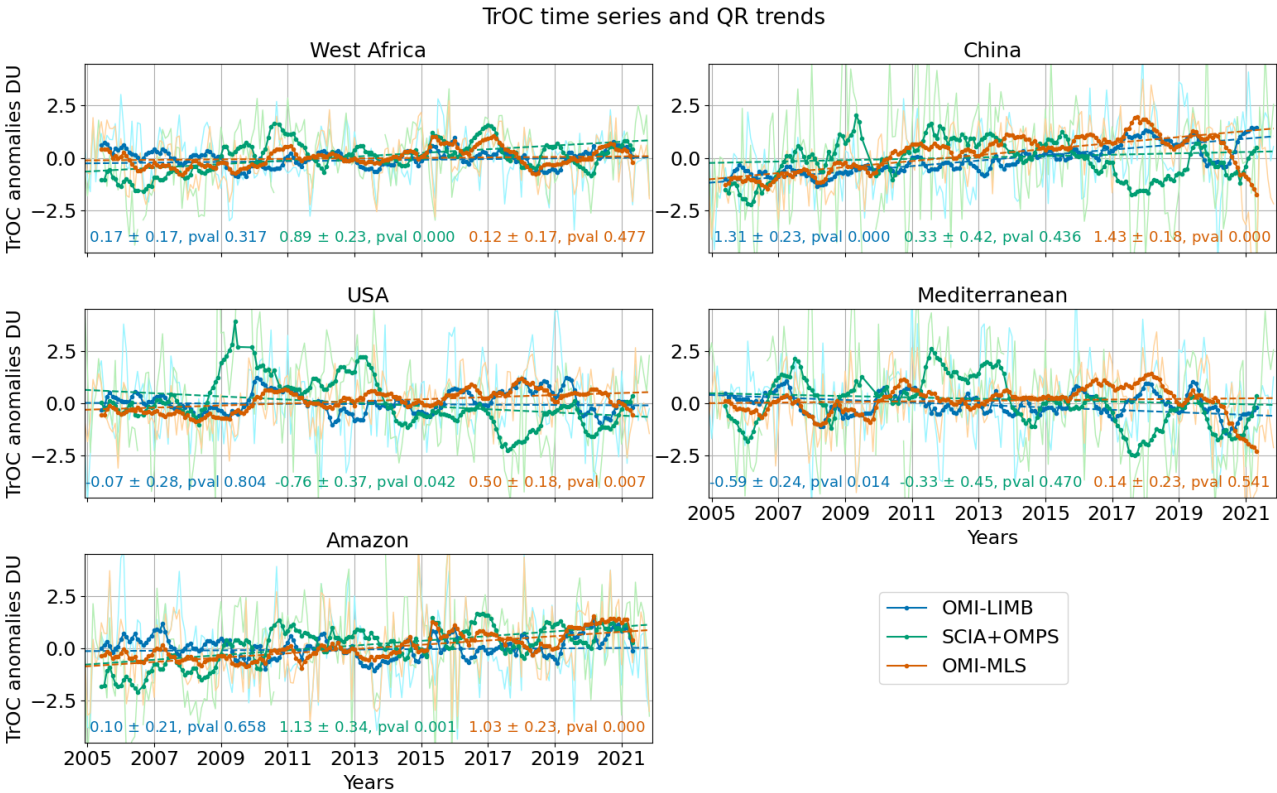
Figure 10. TrOC time series in terms of absolute anomalies (with QBO and ENSO fit contributions subtracted) averaged in specific regions for three datasets over the period ~~2004-2022~~2005-2021. Linear fits are over-plotted and trends reported in DU per decade.

360

---

[1] The conversion DU to ppb was done with the following approximations: $1\ DU = 2.14 * 10^4\ \mu g(O_3)/m^2$, assuming a 10 km layer and $1\ \mu g/m^3 = 0.5$ ppbv

Table 2. Trend values in ppb per decade in the five defined regions, over the period 2005-2021, for three datasets, with 2σ uncertainties, p-values, and respective trend confidence. Recommendations in Chang et al. (2023b) were followed.

Region	Data set	Trend $\pm 2\sigma$ [ppb/dec]	p-value	confidence
West Africa	OMI-LIMB	0.73 ± 0.73	0.31	low
	SCIA+OMPS	3.81 ± 0.99	<0.01	high
	OMI-MLS	0.51 ± 0.73	>0.33	low
China	OMI-LIMB	5.62 ± 0.98	<0.01	high
	SCIA+OMPS	1.41 ± 1.80	>0.33	low
	OMI-MLS	6.13 ± 0.77	<0.01	high
USA	OMI-LIMB	-0.30 ± 1.20	>0.33	low
	SCIA+OMPS	-3.26 ± 1.59	0.04	medium
	OMI-MLS	2.14 ± 0.77	<0.01	high
Mediterranean	OMI-LIMB	-2.53 ± 1.03	0.01	high
	SCIA+OMPS	-1.42 ± 1.93	>0.33	low
	OMI-MLS	0.60 ± 0.99	>0.33	low
Amazon	OMI-LIMB	-0.43 ± 0.90	>0.33	low
	SCIA+OMPS	4.84 ± 1.46	<0.01	high
	OMI-MLS	4.41 ± 0.99	<0.01	high

The only region with a ~~clear and significant~~ positive trend in all datasets is ~~, as expected, Southeast Asia with positive trends up to~~ China with values up to +1.5 DU per decade for OMI-MLS (p-value < 0.01, i.e. very high confidence) and closer to zero for ~~OMI-LIMB. The positive trend from~~ SCIA+OMPS (p-value > 0.33, i.e. very low confidence), with a hint of a possible change in trend sign from 2017. This is consistent with Froidevaux et al. (2025); Gaudel et al. (2024) and in agreement with Lu et al. (2024) using ozonesondes and IAGOS (In-service Aircraft for a Global Observing System) data. However, we need to take into consideration the presence of a positive TPH-related trend, described in Fig. 7, which reduces the confidence of the found positive trends. Positive trends are also seen by SCIA+OMPS and OMPS-MLS in the Amazon ~~is likely related to artifacts in the datasets at the very beginning and end of the time period~~but are not confirmed by OMI-LIMB. OMI-MLS also shows a positive drift with respect to sondes in the tropics, which gives less confidence to this value. Close to zero trends are found in the West Africa region and the US (except for SCIA+OMPS). Summer-time trends over the US are also close to zero (not shown). The Mediterranean region shows ~~significant negative trends,~~negative or close to zero trends for OMI-MLS, but with p-values > 0.2 (low confidence). These negative trends are possibly related to the EU policies introduced to improve air quality. We further investigated trends in Europe, as shown in Fig. ~~S15 shows trends over the whole Europe~~S8, where negative ~~but mostly non-significant values are observed, and~~ low-confidence trends are observed. In particular, close to zero summer-time changes ~~over Europe and the Mediterranean. Only over the latter region, negative and significant~~ are detected over Europe, whereas, over the Mediterranean, all three datasets show negative summer-time trends~~are shown by all three datasets, although~~

with large p-values. Due to the possible positive TPH-related ozone trend at northern mid-latitudes of about 1 DU per decade from Fig. 7, we are more confident in attributing high confidence to these detected negative trends.

When examining the global distribution of ~~MLR~~ QR trends for the same datasets over the ~~2004-2022~~ 2005-2021 time frame, ~~as shown in Fig. ??, with shadings indicating areas with insignificant trends at the 95% confidence level, we notice~~ we noticed overall larger differences between the datasets, with OMI-MLS ~~and SCIA+OMPS~~ showing mostly positive trends. ~~The overall positive values in the tropics from these two datasets are influenced by the positive drift with respect to ozonesondes identified in Fig. 8. Map of TrOC trends for three datasets using the MLR model, over 2004-2022, in per decade. Shaded areas indicate non-significant trends at 2σ.~~

~~This demonstrates~~ This indicates that focusing on specific geographical regions and understanding the differences and the possible artefacts in the time series is required before we are able to provide a global picture of TrOC trends from satellite datasets.

7 Conclusions

We performed an inter-comparison of several limb-nadir TOR datasets to assess their consistency, with a focus on finding approaches to ~~reconcile the discrepancies between them and~~ look for similarities between them rather than highlight ~~the inherent~~ their differences.

The analysis revealed overall similarities in TrOC patterns across the datasets, such as the typical longitudinal asymmetry in the tropics and the maximum over Southeast Asia. However, notable differences in their seasonality, particularly in the southern hemisphere, have been highlighted. Some datasets show a drop in TrOC levels observed from 2020 onward; this has been discussed by looking at possible discontinuities in the used TPH and ~~in reanalysis data ,~~ reanalysis data without finding a convincing cause for it to be an artifact. Further investigations are required.

The biases due to the differences in TPH definition were found to be minor compared to other dataset-specific discrepancies; therefore, the effort to mitigate TPH-related biases in the level 3 data (gridded data) did not show a consistent reduction in the spread of TrOC values from different datasets. We suggest testing similar corrections in level 2 (daily or single profiles) data. We also found that trends in the TPH time series, used by the various satellite datasets to derive tropospheric ozone columns, ~~generally vary within~~ are different from zero. These trends were also converted in terms of TrOC changes, finding the largest contribution at mid-latitudes of up to +1 DU per decade, which needs to be kept in mind when computing TrOC trends.

Additionally, the drift in TOR products was assessed by collocating and comparing them to HEGIFTOM time series. We performed a latitude band-wise analysis, which revealed the presence of small but significant average drifts for some datasets ~~,~~ ~~within 0.9~~ within 0.6 DU per decade at mid-latitudes (excluding OMPS-MERRA) and up to 1.2 DU per decade in the tropics. These ~~patterns~~ values need to be taken into account when evaluating TrOC trends, ~~and~~ however, a direct subtraction of these drift values from TrOC trends is not straightforward. Satellite datasets in the tropics fail to capture short-term features shown by sonde data, which are most probably coming from the lowermost troposphere: a better characterization of the identified patterns is relevant for assessing the significance of TrOC trends. This analysis shows the value of satellite data providing

410 global coverage ~~, and~~and, at the same time~~the need of ~~, the need for stable long-term observations, e.g. ozonesondes, to assess drifts and discontinuities in their time series.

Our investigation of the TrOC trends over the ~~2004~~2005–~~2022~~2021 period focused on specific regions of interest rather than a global analysis~~, revealing significant~~. We applied a QR model including proxies to three long-period datasets and found consistent trends only in two areas: positive over ~~Southeast Asia~~China and negative over the Mediterranean, although with low confidence for the latter region. In most cases, the detected trends are well within ±1 DU per decade.

Activities to homogenize these datasets in terms of TPH and a thorough comparison of satellite observations with ground-based data, consistently taking into account the different vertical resolutions, are recommended to better understand and reconcile the detected discrepancies between satellite datasets and interpret the trend results. Our analysis shows that ~~the homogenization is a crucial step~~ a clear understanding of drifts and biases is crucial before using the datasets for global trends studies.

Data availability. GTO-LIMB and OMI-LIMB are available for download at https://nsdc.fmi.fi/data/data_sunlit.php (last access: 28.11.2024). OMI-MLS, OMPS-MERRA and EPIC-MERRA are publicly available at https://acd-ext.gsfc.nasa.gov/Data_services/cloud_slice/ (last access: 28.11.2024). The other two datasets are available upon request to the co-authors of this study. HEGIFTOM data are accessible through the following webpage: https://hegiftom.meteo.be/datasets/tropospheric-ozone-columns-trocs (last access: 28.11.2024).

Appendix A

Since the averaging kernels (AK) of TOC retrievals typically show a lower sensitivity to the lowermost troposphere, we tested the role of the close-to-surface ozone in terms of effects on the drift analysis (and trends, see Fig. B2). The application of satellite-specific AK is beyond the scope of this work; however, to take into account the different vertical sensitivity of the TOC measurements with respect to high-resolved ozonesondes, we tested a simple weighting of the sonde profiles before vertical integration, as described by a function providing a rough approximation of OMI averaging kernels:

$$w = \left(\frac{z}{z_{lim}} \right)^{0.5} \text{ for z} < z_{lim} \;, \tag{A1}$$

where z is the sonde altitude, z_{lim} is two thirds of the TPH and the weight w is equal to 1 above this altitude. Approximated OMI AKs are reported in the Supplements, Fig. S9. More examples of OMI AK can be found in e.g. Fig. S3 of Sofieva et al. (2022) . This paper also discusses the changes in TrOC retrieved by the residual method that are caused by the low sensitivity of nadir instruments in the lowermost troposphere.

Some caveats shall be added: the weighting from Eq. A1 does not conserve the TrOC value, so that this approach is suitable to investigate TrOC anomalies only. In addition, we are aware that the air mass factor in TOC retrieval takes into account the reduced sensitivity of the lowermost troposphere. This however might not be sufficient to capture cases with ozone

concentrations in the boundary layer far from climatological values or cases when the TOC AK gets close to zero in the lowermost troposphere.

The relevant outcome of this weighting for the present manuscript is related to the drift of the satellite data in the tropics, shown in Fig. A1. In particular, by reducing the lowermost tropospheric contribution from the ozonesondes, drift values in the tropics are reduced (except for SCIA+OMPS). In addition, and most importantly, in comparison to panel (b) of Fig.8, satellite data better capture short-period patterns present in sonde data. In fact, the HEGIFTOM anomaly line (black) is less correlated to the TOR-HEGIFTOM residuals, indicating that reducing the weight of close-to-surface ozone improves the agreement with satellite data. This conclusion needs some further investigations.
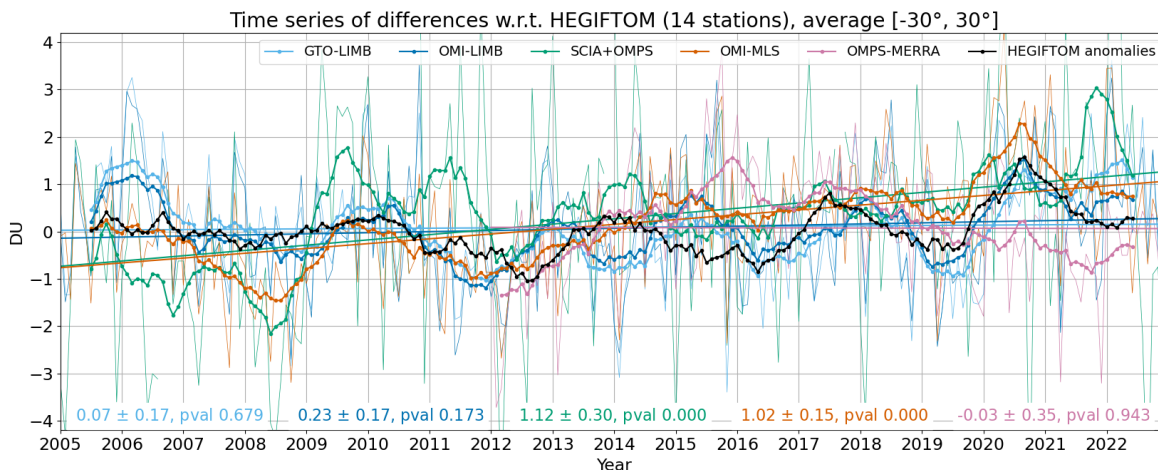
Figure A1. Drift in DU/dec of the difference time series between satellite and HEGIFTOM sonde anomalies, after the weighting in Eq.A1, in the tropics. Thin lines indicate the monthly difference time series, whereas thicker lines are their 13-month averages; the HEGIFTOM anomalies are also shown (with sign changed). Linear trend values (drifts) with corresponding 2σ uncertainties and p-values are reported at the bottom of the panels. The titles include the number of stations used for the analysis.

Appendix B

An overview of the comparison between ozonesonde data and the TOR products is given in Figs. B1 and B2, with the aim of providing an idea of the overall agreement between the datasets.

Figure B1 shows the relationship between the relative mean bias of each satellite dataset with respect to HEGIFTOM data (without AK application) on the y-axis and the respective standard deviation of the relative differences on the x-axis. Each dot is a sonde station, and the color corresponds to its latitude. We notice the negative bias of the first two datasets with respect to most of the sonde TrOC values. On the contrary, S5P-BASCOE generally shows a high bias, particularly at northern mid-and high latitudes. The standard deviation of the differences tends to increase with latitude for most products, except for OMI-MLS,

455   OMPS-MERRA and EPIC-MERRA. On the contrary, there is no indication of a dependence of the mean bias from latitude. We also notice the generally higher variability characterizing the SCIA+OMPS dataset.
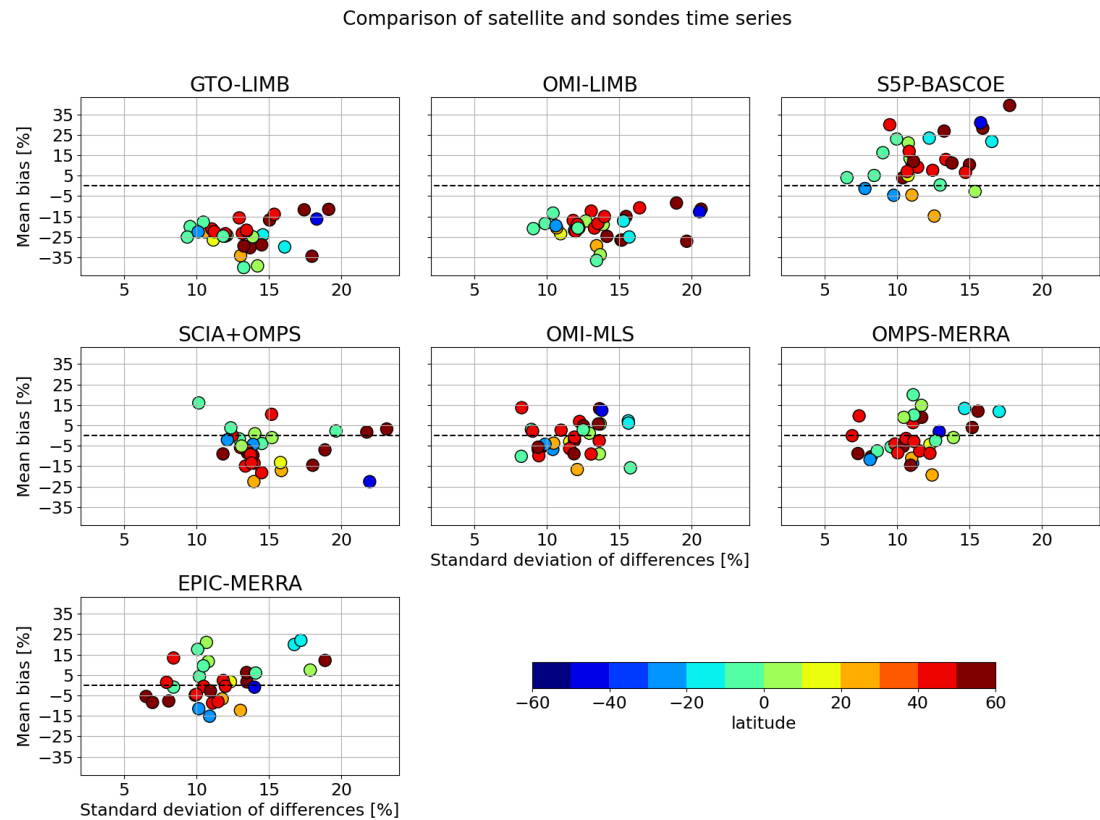
Figure B1. Scatter plot of the mean bias between HEGIFTOM data and each satellite dataset against the respective standard deviation of the differences, in percentage values. Each dot is a sonde station and the color corresponds to its latitude.

    Figure B2 displays the trend values of the HEGIFTOM time series and of the collocated TOR products, in percentage per decade. These were computed by applying the ~~MLR~~ QR model described in Sect. 6 to the relative anomalies of each time series. Only the three satellite products with the longest time series have been considered, taking into account that GTO-LIMB

460   shows very similar results to OMI-LIMB so that it is not displayed here. For some stations, the trends do not show a good agreement between sondes and satellite data or even among the satellite products. We also include the HEGIFTOM trends computed applying the AK weighting: these values generally show smaller discrepancies with respect to the satellite products and closer to zero trends overall.

Appendix C
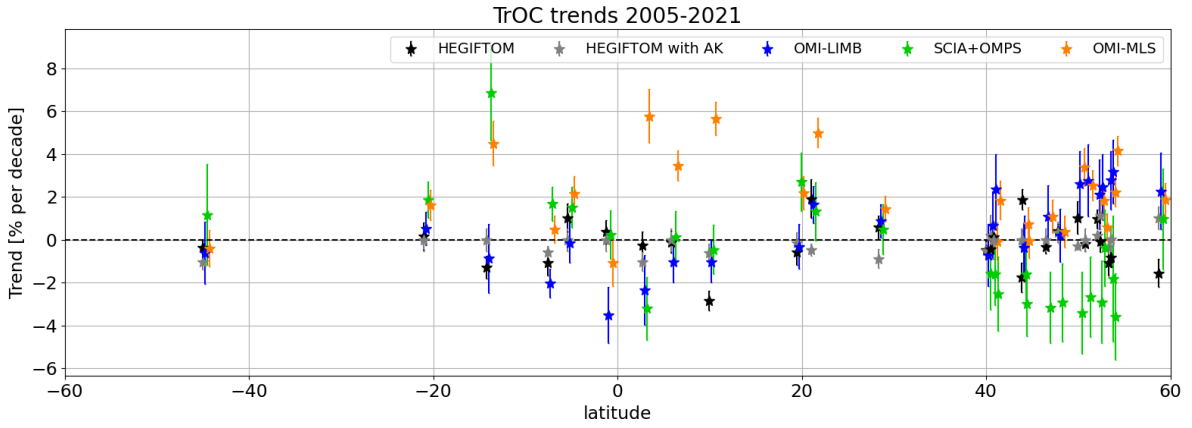
Figure B2. Trend values of the HEGIFTOM time series and of the collocated TOR datasets, in percentage per decade over 2005-2021. Only the three products with the longest time series are displayed (taking into account that GTO-LIMB results are very similar to OMI-LIMB)

465 ~~To better understand the behaviour of the three regression models at different locations, Fig. **??** shows a comparison of the results applying three regression models to the OMI-LIMB dataset: a linear fit (LIN), the quantile regression (QR) and a multi-linear fit (MLR) including QBO and ENSO proxies. The considered period is from 2005 to 2022 and trends are in DU per decade. The top panels show the results for LIN and QR, whereas the bottom panels show the difference in trends with respect to MLR results (we do not include here the MLR map, which can be found in Fig. **??**). Uncertainties are not~~
470 ~~over-plotted to better visualize the differences in the trend values between models. We notice that the largest differences (note the different color scale) between LIN and MLR are located in the ENSO region. The differences between MLR and QR show a more random pattern, with overall increasing values towards high latitudes, related to the largest scatter of the values or to the presence of outliers.~~

~~Comparison of the trend results (DU per decade) from three regression models applied to OMI-LIMB data over 2005-2021.~~
475 ~~Top panel: trends from LIN and QR; bottom panel: differences of trend values (MLR - LIN) and (MLR - QR).~~

Author contributions. CA performed most of the analysis for the inter-comparison of the datasets and wrote the manuscript. VS provided guidance and expertise for the analysis strategy and delivered the OMI-LIMB and GTO-LIMB datasets. AOC provided the SCIA+OMPS dataset and supported the work on the TPH correction. AR provided the expertise on limb-nadir matching and supervised the data analysis. KPH and DL delivered the S5P-BASCOE dataset. ED supervised the ESA-related OREGANO project. DT and RMS gave inputs and suggestions regarding the comparison with sondes. RVM gave support regarding the usage of the HEGIFTOM dataset. JZ delivered OMI-MLS, OMPS-MERRA and EPIC-MERRA and gave insights into the TrOC drop in recent years. MW, who leads the project, contributed to the scientific outcome. All co-authors contributed to the review of the manuscript.

Competing interests. Some co-authors are members of the editorial board of AMT.

Acknowledgements. This study has been primarily funded by ESA within the OREGANO (Ozone Recovery from Merged Observational Data and Model Analysis) project, and by the State and University of Bremen. GTO-LIMB and OMI-LIMB have been developed within the ESA CCI (Climate Change Initiative) project. We also acknowledge ESA for funding a six-month research stay for CA at the ESRIN Science Hub. Some of the ozonesonde data used in this publication are part of the Network for the Detection of Atmospheric Composition Change (NDACC) and are available through the NDACC website www.ndacc.org. We also acknowledge the data providers within SHADOZ and WOUDC networks. The processing of OMPS and SCIAMACHY stratospheric ozone and aerosol profiles has been done ~~on the German HLRN (High Performance Computer Center North~~at the German NHR (National High Performance Computing) alliance, within the projects number hbk00063 and hbk00092. Part of the processing was also done on the Hypatia HPC facility at IUP, funded under DFG/FUGG grant nos. INST 144/379-1 and INST 144/493-1. The GALAHAD library was used for the processing. VS thanks the Academy of Finland (Centre of Excellence of Inverse Modelling and Imaging; decision no. 353082). We would like to thank Daan Hubert for the fruitful exchange of ideas and analysis approaches. We also acknowledge the contribution of the reviewers that helped improving the manuscript and the valuable advices from Owen Cooper.

References

Bak, J., Liu, X., Wei, J., Pan, L., Chance, K., and Kim, J.: Improvement of OMI ozone profile retrievals in the upper troposphere and lower stratosphere by the use of a tropopause-based ozone profile climatology, Atmospheric Measurement Techniques, 6, 2239–2254, 2013.

Boynard, A., Clerbaux, C., Coheur, P.-F., Hurtmans, D., Turquety, S., George, M., Hadji-Lazaro, J., Keim, C., and Meyer-Arnek, J.: Measurements of total and tropospheric ozone from IASI: comparison with correlative satellite, ground-based and ozonesonde observations, Atmospheric Chemistry and Physics, 9, 6255–6271, 2009.

Brown, J., Bowman, C., et al.: Integrated science assessment for ozone and related photochemical oxidants, US Environmental Protection Agency, 2013.

Bruckner, M., Pierce, R. B., and Lenzen, A.: Examining ENSO-related variability in tropical tropospheric ozone in the RAQMS-Aura chemical reanalysis, Atmospheric Chemistry and Physics, 24, 10 921–10 945, 2024.

Chang, K.-L., Cooper, O. R., Gaudel, A., Allaart, M., Ancellet, G., Clark, H., Godin-Beekmann, S., Leblanc, T., Van Malderen, R., Nédélec, P., et al.: Impact of the COVID-19 economic downturn on tropospheric ozone trends: An uncertainty weighted data synthesis for quantifying regional anomalies above western North America and Europe, AGU Advances, 3, e2021AV000 542, 2022.

Chang, K.-L., Cooper, O. R., Rodriguez, G., Iraci, L. T., Yates, E. L., Johnson, M. S., Gaudel, A., Jaffe, D. A., Bernays, N., Clark, H., et al.: Diverging ozone trends above western North America: Boundary layer decreases versus free tropospheric increases, Journal of Geophysical Research: Atmospheres, 128, e2022JD038 090, 2023a.

Chang, K.-L., Schultz, M. G., Koren, G., and Selke, N.: Guidance note on best statistical practices for TOAR analyses, arXiv preprint arXiv:2304.14236, 2023b.

Christiansen, A., Mickley, L. J., Liu, J., Oman, L. D., and Hu, L.: Multidecadal increases in global tropospheric ozone derived from ozonesonde and surface site observations: can models reproduce ozone trends?, Atmospheric Chemistry and Physics, 22, 14 751–14 782, 2022.

Coldewey-Egbers, M., Loyola, D. G., Lerot, C., and Van Roozendael, M.: Global, regional and seasonal analysis of total ozone trends derived from the 1995–2020 GTO-ECV climate data record, Atmospheric Chemistry and Physics, 22, 6861–6878, 2022.

Ebojie, F., Burrows, J., Gebhardt, C., Ladstätter-Weißenmayer, A., Von Savigny, C., Rozanov, A., Weber, M., and Bovensmann, H.: Global tropospheric ozone variations from 2003 to 2011 as seen by SCIAMACHY, Atmospheric Chemistry and Physics, 16, 417–436, 2016.

Elshorbany, Y., Ziemke, J. R., Strode, S., Petetin, H., Miyazaki, K., De Smedt, I., Pickering, K., Seguel, R. J., Worden, H., Emmerichs, T., et al.: Tropospheric ozone precursors: global and regional distributions, trends, and variability, Atmospheric Chemistry and Physics, 24, 12 225–12 257, 2024.

Fishman, J. and Larsen, J. C.: Distribution of total ozone and stratospheric ozone in the tropics: Implications for the distribution of tropospheric ozone, Journal of Geophysical Research: Atmospheres, 92, 6627–6634, 1987.

Fishman, J., Watson, C. E., Larsen, J. C., and Logan, J. A.: Distribution of tropospheric ozone determined from satellite data, Journal of Geophysical Research: Atmospheres, 95, 3599–3617, 1990.

Fleming, Z. L., Doherty, R. M., Von Schneidemesser, E., Malley, C. S., Cooper, O. R., Pinto, J. P., Colette, A., Xu, X., Simpson, D., Schultz, M. G., et al.: Tropospheric Ozone Assessment Report: Present-day ozone distribution and trends relevant to human health, Elem Sci Anth, 6, 12, 2018.

Froidevaux, L., Kinnison, D. E., Gaubert, B., Schwartz, M. J., Livesey, N. J., Read, W. G., Bardeen, C. G., Ziemke, J. R., and Fuller, R. A.: Tropical upper-tropospheric trends in ozone and carbon monoxide (2005–2020): observational and model results, Atmospheric Chemistry and Physics, 25, 597–624, 2025.

535 Garane, K., Koukouli, M.-E., Verhoelst, T., Lerot, C., Heue, K.-P., Fioletov, V., Balis, D., Bais, A., Bazureau, A., Dehn, A., et al.: TROPOMI/S5P total ozone column data: global ground-based validation and consistency with other satellite missions, Atmospheric Measurement Techniques, 12, 5263–5287, 2019.

Gaudel, A., Cooper, O. R., Ancellet, G., Barret, B., Boynard, A., Burrows, J. P., Clerbaux, C., Coheur, P.-F., Cuesta, J., Cuevas, E., et al.: Tropospheric Ozone Assessment Report: Present-day distribution and trends of tropospheric ozone relevant to climate and global atmospheric chemistry model evaluation, Elem Sci Anth, 6, 39, 2018.

540 Gaudel, A., Bourgeois, I., Li, M., Chang, K.-L., Ziemke, J., Sauvage, B., Stauffer, R. M., Thompson, A. M., Kollonige, D. E., Smith, N., et al.: Tropical tropospheric ozone distribution and trends from in situ and satellite data, Atmospheric Chemistry and Physics, 24, 9975–10 000, 2024.

Gelaro, R., McCarty, W., Suárez, M. J., Todling, R., Molod, A., Takacs, L., Randles, C. A., Darmenov, A., Bosilovich, M. G., Reichle, R., et al.: The modern-era retrospective analysis for research and applications, version 2 (MERRA-2), Journal of climate, 30, 5419–5454, 545 2017.

Heue, K.-P., Loyola, D., Romahn, F., Zimmer, W., Chabrillat, S., Errera, Q., Ziemke, J., and Kramarova, N.: Tropospheric ozone retrieval by a combination of TROPOMI/S5P measurements with BASCOE assimilated data, Atmospheric Measurement Techniques, 15, 5563–5579, 2022.

Keppens, A., Di Pede, S., Hubert, D., Lambert, J.-C., Veefkind, P., Sneep, M., De Haan, J., ter Linden, M., Leblanc, T., Compernolle, S., 550 et al.: Five years of Sentinel-5p TROPOMI operational ozone profiling and geophysical validation using ozonesonde and lidar ground-based networks, Atmospheric Measurement Techniques Discussions, 2024, 1–43, 2024.

Lu, X., Liu, Y., Su, J., Weng, X., Ansari, T., Zhang, Y., He, G., Zhu, Y., Wang, H., Zeng, G., et al.: Tropospheric ozone trends and attributions over East and Southeast Asia in 1995–2019: An integrated assessment using statistical methods, machine learning models, and multiple chemical transport models, EGUsphere, 2024, 1–65, 2024.

555 Marshak, A., Herman, J., Adam, S., Karin, B., Carn, S., Cede, A., Geogdzhayev, I., Huang, D., Huang, L.-K., Knyazikhin, Y., et al.: Earth observations from DSCOVR EPIC instrument, Bulletin of the American Meteorological Society, 99, 1829–1850, 2018.

Mettig, N., Weber, M., Rozanov, A., Arosio, C., Burrows, J. P., Veefkind, P., Thompson, A. M., Querel, R., Leblanc, T., Godin-Beekmann, S., et al.: Ozone profile retrieval from nadir TROPOMI measurements in the UV range, Atmospheric Measurement Techniques, 14, 6057–6082, 2021.

560 Miles, G., Siddans, R., Kerridge, B., Latter, B., and Richards, N.: Tropospheric ozone and ozone profiles retrieved from GOME-2 and their validation, Atmospheric Measurement Techniques, 8, 385–398, 2015.

Mills, G., Pleijel, H., Malley, C. S., Sinha, B., Cooper, O. R., Schultz, M. G., Neufeld, H. S., Simpson, D., Sharps, K., Feng, Z., et al.: Tropospheric Ozone Assessment Report: Present-day tropospheric ozone distribution and trends relevant to vegetation, Elem Sci Anth, 6, 47, 2018.

565 Miyazaki, K., Bowman, K., Sekiya, T., Takigawa, M., Neu, J. L., Sudo, K., Osterman, G., and Eskes, H.: Global tropospheric ozone responses to reduced NO x emissions linked to the COVID-19 worldwide lockdowns, Science Advances, 7, eabf7460, 2021.

Monks, P. S., Archibald, A., Colette, A., Cooper, O., Coyle, M., Derwent, R., Fowler, D., Granier, C., Law, K. S., Mills, G., et al.: Tropospheric ozone and its precursors from the urban to the global scale from air quality to short-lived climate forcer, Atmospheric chemistry and physics, 15, 8889–8973, 2015.

Monsees, F., Rozanov, A., Burrows, J. P., Weber, M., Rinke, A., Jaiser, R., and von der Gathen, P.: Relations between cyclones and ozone changes in the Arctic using data from satellite instruments and the MOSAiC ship campaign, Atmospheric Chemistry and Physics, 24, 9085–9099, 2024.

Munro, R., Siddans, R., Reburn, W. J., and Kerridge, B. J.: Direct measurement of tropospheric ozone distributions from space, Nature, 392, 168–171, 1998.

Orfanoz-Cheuquelaf, A., Arosio, C., Rozanov, A., Weber, M., Ladstätter-Weißenmayer, A., Burrows, J. P., Thompson, A. M., Stauffer, R. M., and Kollonige, D. E.: Tropospheric ozone column dataset from OMPS-LP/OMPS-NM limb-nadir matching, Atmospheric Measurement Techniques Discussions, 2023, 1–28, 2023.

Pimlott, M. A., Pope, R. J., Kerridge, B. J., Siddans, R., Latter, B. G., Ventress, L. J., Feng, W., and Chipperfield, M. P.: Large Reductions in Satellite-Derived and Modelled European Lower Tropospheric Ozone During and After the COVID-19 Pandemic (2020–2022), EGUsphere, 2024, 1–16, 2024.

Pope, R. J., Kerridge, B. J., Siddans, R., Latter, B. G., Chipperfield, M. P., Feng, W., Pimlott, M. A., Dhomse, S. S., Retscher, C., and Rigby, R.: Investigation of spatial and temporal variability in lower tropospheric ozone from RAL Space UV–Vis satellite products, Atmospheric Chemistry and Physics, 23, 14 933–14 947, 2023.

Pope, R. J., O'Connor, F. M., Dalvi, M., Kerridge, B. J., Siddans, R., Latter, B. G., Barret, B., Le Flochmoen, E., Boynard, A., Chipperfield, M. P., et al.: Investigation of the impact of satellite vertical sensitivity on long-term retrieved lower-tropospheric ozone trends, Atmospheric Chemistry and Physics, 24, 9177–9195, 2024.

Putero, D., Cristofanelli, P., Chang, K.-L., Dufour, G., Beachley, G., Couret, C., Effertz, P., Jaffe, D. A., Kubistin, D., Lynch, J., et al.: Fingerprints of the COVID-19 economic downturn and recovery on ozone anomalies at high-elevation sites in North America and western Europe, Atmospheric Chemistry and Physics, 23, 15 693–15 709, 2023.

Reichler, T., Dameris, M., and Sausen, R.: Determining the tropopause height from gridded data, Geophysical research letters, 30, 2003.

Rowlinson, M. J., Rap, A., Arnold, S. R., Pope, R. J., Chipperfield, M. P., McNorton, J., Forster, P., Gordon, H., Pringle, K. J., Feng, W., et al.: Impact of El Niño–Southern Oscillation on the interannual variability of methane and tropospheric ozone, Atmospheric Chemistry and Physics, 19, 8669–8686, 2019.

Sofieva, V. F., Hänninen, R., Sofiev, M., Szeląg, M., Lee, H. S., Tamminen, J., and Retscher, C.: Synergy of using nadir and limb instruments for tropospheric ozone monitoring (SUNLIT), Atmospheric Measurement Techniques, 15, 3193–3212, 2022.

Stauffer, R. M., Thompson, A. M., Kollonige, D. E., Tarasick, D. W., Van Malderen, R., Smit, H. G., Vömel, H., Morris, G. A., Johnson, B. J., Cullis, P. D., et al.: An examination of the recent stability of ozonesonde global network data, Earth and Space Science, 9, e2022EA002 459, 2022.

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., et al.: Dynamical drivers of free-tropospheric ozone increases over equatorial Southeast Asia, Atmospheric Chemistry and Physics, 24, 5221–5234, 2024.

Steinbrecht, W., Kubistin, D., Plass-Dülmer, C., Davies, J., Tarasick, D. W., von Der Gathen, P., Deckelmann, H., Jepsen, N., Kivi, R., Lyall, N., et al.: COVID-19 crisis reduces free tropospheric ozone across the Northern Hemisphere, Geophysical Research Letters, 48, e2020GL091 987, 2021.

605 Tarasick, D., Galbally, I. E., Cooper, O. R., Schultz, M. G., Ancellet, G., Leblanc, T., Wallington, T. J., Ziemke, J., Liu, X., Steinbacher, M., et al.: Tropospheric Ozone Assessment Report: Tropospheric ozone from 1877 to 2016, observed levels, trends and uncertainties, Elem Sci Anth, 7, 39, 2019.

Van Malderen, R. and Smit, H. G. J.: HEGIFTOM webpage, https://hegiftom.meteo.be/, accessed: 2024-11-21, 2020.

Van Malderen, R., Thompson, A. M., Kollonige, D. E., Stauffer, R. M., Smit, H. G., Barras, E. M., Vigouroux, C., Petropavlovskikh, I.,
610 Leblanc, T., Thouret, V., 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.

Wang, H., Lu, X., Jacob, D. J., Cooper, O. R., Chang, K.-L., Li, K., Gao, M., Liu, Y., Sheng, B., Wu, K., et al.: Global tropospheric ozone trends, attributions, and radiative impacts in 1995–2017: an integrated analysis using aircraft (IAGOS) observations, ozonesonde, and multi-decadal chemical model simulations, Atmospheric Chemistry and Physics, 22, 13 753–13 782, 2022.

615 Wargan, K., Labow, G., Frith, S., Pawson, S., Livesey, N., and Partyka, G.: Evaluation of the Ozone Fields in NASA's MERRA-2 Reanalysis, Journal of Climate, 30, 2961–2988, 2017.

Warneck, P.: Chemistry of the natural atmosphere, vol. 71, Elsevier, 1999.

WMO: Meteorology a three-dimensional science: second session of the commission for aerology, WMO Bulletin IV (4), 1957.

Zhang, L., Jacob, D. J., Liu, X., Logan, J. A., Chance, K., Eldering, A., and Bojkov, B. R.: Intercomparison methods for satellite mea-
620 surements of atmospheric composition: application to tropospheric ozone from TES and OMI, Atmospheric Chemistry and Physics, 10, 4725–4739, 2010.

Ziemke, J., Chandra, S., and Bhartia, P.: Two new methods for deriving tropospheric column ozone from TOMS measurements: Assimilated UARS MLS/HALOE and convective-cloud differential techniques, Journal of Geophysical Research: Atmospheres, 103, 22 115–22 127, 1998.

625 Ziemke, J., Chandra, S., and Bhartia, P.: "Cloud slicing": A new technique to derive upper tropospheric ozone from satellite measurements, Journal of Geophysical Research: Atmospheres, 106, 9853–9867, 2001.

Ziemke, J., Chandra, S., Duncan, B., Froidevaux, L., Bhartia, P., Levelt, P., and Waters, J.: Tropospheric ozone determined from Aura OMI and MLS: Evaluation of measurements and comparison with the Global Modeling Initiative's Chemical Transport Model, Journal of Geophysical Research: Atmospheres, 111, 2006.

630 Ziemke, J. R., Oman, L. D., Strode, S. A., Douglass, A. R., Olsen, M. A., McPeters, R. D., Bhartia, P. K., Froidevaux, L., Labow, G. J., Witte, J. C., et al.: Trends in global tropospheric ozone inferred from a composite record of TOMS/OMI/MLS/OMPS satellite measurements and the MERRA-2 GMI simulation, Atmospheric Chemistry and Physics, 19, 3257–3269, 2019.

Ziemke, J. R., Kramarova, N. A., Frith, S. M., Huang, L.-K., Haffner, D. P., Wargan, K., Lamsal, L. N., Labow, G. J., McPeters, R. D., and Bhartia, P. K.: NASA Satellite Measurements Show Global-Scale Reductions in Free Tropospheric Ozone in 2020 and Again in 2021
635 During COVID-19, Geophysical Research Letters, 49, e2022GL098 712, 2022.