Response to Referee #1 comment on "Global Ground-based Tropospheric Ozone Measurements: Reference Data and Individual Site Trends (2000–2022) from the TOAR-II/HEGIFTOM Project" by Van Malderen et al.

This manuscript uses homogenized ozone data from ozonesondes, IAGOS, FTIR, Lidar, and Umkehr instruments to analyze global trends in total, free, and lower tropospheric ozone. The authors compared different regressions and explored reasons for trend differences among stations with co-located instruments. Overall, this manuscript is excellent, and these ozone datasets and analyses are critically needed in the ozone community. I appreciate the authors' careful consideration and handling of sparse and complex ozone data. I have only minor comments.

Thank you very much for your positive feedback and for taking your time to review the manuscript. We performed new analyses as explained and illustrated below. In several cases, more material was added to the Supplement. The main conclusions of the study, expressed in the Abstract and last Section of the manuscript have not changed substantially.

Page 2, Line 72-73: There is a reference to "Ref to Elementa collection". Please fix this so it is a real, traceable citation.

This can be dropped, the Gaudel et al. (2018) reference is the only one that has been cited. We also included some references to relevant new TOAR-II papers, e.g. Arosio et al. (2024), Pennington et al. (2024), Jones et al. (2025), Gong et al. (2025), Dufour et al. (2025, Keppens et al. (2025), Maratt Satheesan et al. (2025).

Page 5, Line 154: Are there any results from the ASOPOS WMO GAW report that influenced how the ozonesonde data was handled?

The ASOPOS WMO GAW reports give recommendations on how to prepare, process, and archive ozonesonde data. The second report WMO GAW No. 268 incorporates the O3S-DQA homogenization activity, in the sense that its recommendations on data processing align with those of the O3S-DQA homogenization activity. We tried to make this clearer by modifying this sentence to "Following the GAW Report 201, the most recent JOSIE took place in 2017 (Thompson et al., 2019), which, together with the activity described in the next paragraph, led to a second ASOPOS WMO GAW Report (Report 268, Smit et al., 2021). "And after the description of the O3S-DQA principles used in our study, we now added the sentence "The O3S-DQA guidelines for data processing, standards, and uncertainty estimation are now the current recommendations in WMO-GAW No 268 (Smit et al., 2021)."

Page 7, Line 175: I have some concerns about calculating monthly averages from locations where only one or two ozonesonde measurements are available within a given month. Previous literature suggests that somewhere between 3-18 observations per month are needed for accurate and representative time series (e.g., Christiansen et al., 2022; Lu et al., 2019; Chang et al., 2020; Wang et al., 2022). Could the authors perform some kind of short analysis that compares trends from once- or twice-monthly samples to trends derived from more frequently sampled sites? One thought could be to use a site that has many observations each month, then randomly select two observations to use from each month. Would the trends be similar to those derived from the full dataset? At the least, a discussion of the uncertainty involved in using these very sparse ozonesonde datasets is warranted.

Yes, we are aware of the literature suggesting that 3-18 observations per month are needed for calculating accurate tropospheric ozone trends. The impact of monthly sampling frequency on the trends has already been explicitly discussed in section 4.3.1, in which trends are compared between different techniques at collocated and nearby sites. We followed your guidance on a short analysis and subsampled all the time series of the 55 sites of our sample to two random observations a month, before calculating the monthly means (L3). The QR and MLR L3 trends estimated from this subsampled dataset can then be directly compared with the original QR and MLR L3 trend estimates.

For illustration, we include here a figure (Fig. R1) that shows the L3 monthly mean time series since 2000 for 5 sites (each for one technique) with high monthly sampling frequency (see also Table R1), together with the L3 monthly

mean time series obtained from exactly two randomly chosen measurements a month. It should be clear that reducing the monthly sampling frequency increases the variability of the monthly mean values.

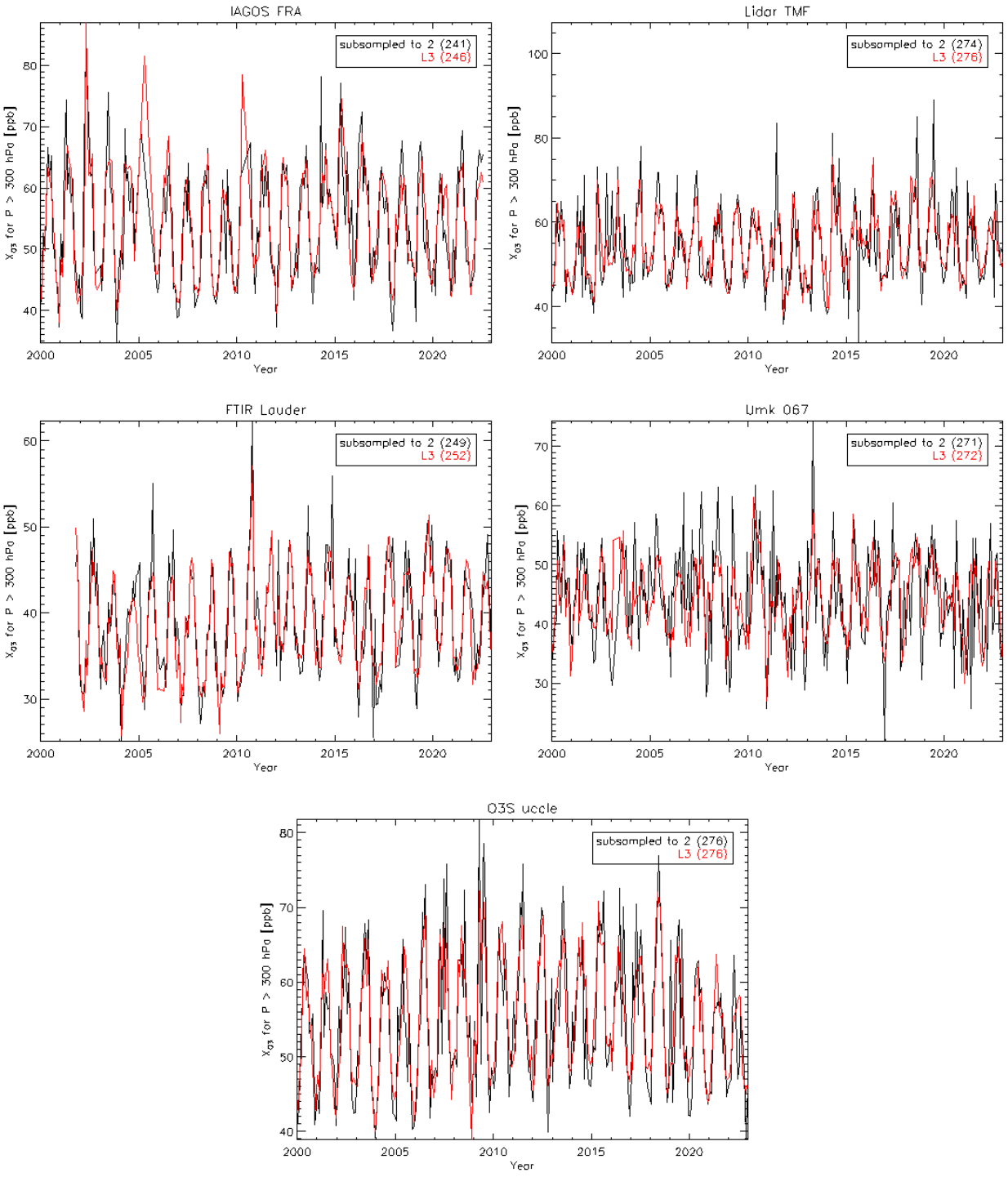

Fig. R1: Monthly mean time series since 2000 using all available measurements (red) and using exactly two randomly chosen daily measurements (black) at Frankfurt airport (IAGOS), Table Mountain Facility (Lidar), Lauder (FTIR), Boulder (Umkehr) and Uccle (ozonesondes). The numbers between brackets in the legend denote the number of months with available data.

Table R1: TrOC QR and MLR L3 trend calculations for the original and subsampled (2 randomly chosen daily measurements a month, denoted by "Sub" in the column headers) monthly mean time series. Trends are estimated in ppb/dec, and trend uncertainties (2 standard deviations) and p-values (QR) are given. Bold trends have p< 0.05. Left columns display the station name, instrument type, and mean monthly sampling frequency of the original time series.

Station	Instrument	Monthly Sampling	QR L3 Trend ± 2*σ (ppb/dec)	QR L3 p-value	QR L3 Sub Trend ± 2*σ (ppb/dec)	QR L3 Sub p-value	MLR L3 Trend ± 2*σ (ppb/dec)	MLR L3 Sub Trend ± 2*σ (ppb/dec)
Northern Hemisphere (180W-20W) TrOC (surface to 300hPa) Trends								
Alert	O3S	4.04	0.74 ± 1.76	0.42	-0.07 ± 2.26	0.95	0.62 ± 1.63	0.42 ± 1.89
ATL	IAGOS	5.51	-0.78 ± 2.22	0.50	N/A	N/A	-0.53 ± 2.44	N/A
Boulder	O3S	4.51	**-1.41 ± 1.14**	**0.01**	-0.17 ± 1.43	0.81	**-1.30 ± 0.79**	-0.91 ± 1.02
Boulder (067)	Umkehr	13.47	0.44 ± 1.30	0.52	-0.82 ± 1.43	0.26	-0.02 ± 1.08	-0.99 ± 1.41
Churchill	O3S	3.63	-1.64 ± 2.42	0.17	-2.85 ± 3.02	0.08	**-3.01 ± 1.98**	-3.02 ± 2.27
DAL	IAGOS	3.08	**2.41 ± 1.66**	**0.01**	N/A	N/A	2.16 ± 2.63	N/A
Edmonton	O3S	3.93	0.03 ± 0.96	0.95	-0.11 ± 1.47	0.89	-0.64 ± 0.95	-0.80 ± 1.11
Eureka	O3S	5.40	0.32 ± 1.36	0.65	-0.03 ± 1.89	0.98	-0.30 ± 1.37	-0.16 ± 1.51
Fairbanks (105)	Umkehr	8.72	0.02 ± 2.28	0.99	-0.15 ± 3.70	0.94	0.98 ± 2.77	0.84 ± 3.72
Goose Bay	O3S	4.09	-0.80 ± 1.28	0.21	-0.80 ± 1.49	0.28	-0.26 ± 1.20	-0.47 ± 1.34
Hilo	O3S	4.13	-0.43 ± 1.30	0.50	-0.86 ± 1.71	0.32	-0.41 ± 1.03	-1.06 ± 1.32
Mauna Loa	FTIR	14.37	1.26 ± 2.48	0.30	0.99 ± 2.54	0.44	0.88 ± 2.33	1.15 ± 3.37
Mauna Loa (031)	Umkehr	19.65	**1.62 ± 0.96**	**0.00**	**1.73 ± 1.12**	**0.00**	**1.49 ± 0.91**	**1.50 ± 1.17**
Paramaribo	O3S	3.40	-0.42 ± 1.04	0.45	-0.16 ± 1.18	0.78	0.22 ± 1.17	-0.10 ± 1.36
Resolute	O3S	3.71	**-2.07 ± 1.78**	**0.03**	-0.80 ± 1.80	0.37	**-2.12 ± 1.80**	-1.63 ± 1.87
Scoresbysund	O3S	4.23	**-2.73 ± 1.40**	**0.00**	**-2.51 ± 1.72**	**0.00**	**-2.82 ± 1.15**	**-3.23 ± 1.38**
TMF	Lidar	10.57	**1.24 ± 1.08**	**0.02**	**1.28 ± 1.08**	**0.02**	**1.31 ± 1.02**	1.08 ± 1.36
Thule	FTIR	9.69	**-3.27 ± 1.74**	**0.00**	-1.77 ± 1.89	0.06	**-3.59 ± 1.92**	-1.41 ± 2.75
Toronto	FTIR	8.98	-1.15 ± 2.16	0.27	-0.72 ± 3.50	0.68	-1.70 ± 2.08	-1.02 ± 3.02
Trinidad Head	O3S	4.50	-0.96 ± 1.12	0.07	**-1.41 ± 1.27**	**0.03**	**-0.90 ± 0.89**	-1.08 ± 1.14
Wallops Island	O3S	4.67	**-2.83 ± 1.50**	**0.00**	**-2.82 ± 1.74**	**0.00**	**-2.81 ± 1.25**	**-2.45 ± 1.42**
Northern Hemisphere (19W-79E) TrOC (surface to 300hPa) Trends								
Arosa (035)	Umkehr	8.09	0.63 ± 1.36	0.34	0.03 ± 2.00	0.98	0.68 ± 1.05	-0.38 ± 1.42
Ascension Island	O3S	3.83	-1.06 ± 1.76	0.22	-0.70 ± 1.54	0.37	-0.88 ± 1.74	-0.93 ± 1.84
De Bilt	O3S	4.31	**1.50 ± 1.20**	**0.01**	**2.08 ± 1.38**	**0.00**	**1.34 ± 1.08**	**1.34 ± 1.34**
FRA	IAGOS	24.64	0.09 ± 1.10	0.87	1.13 ± 1.19	0.06	-0.04 ± 1.08	0.28 ± 1.36
Hohenpeissenberg	O3S	10.58	0.55 ± 0.94	0.23	0.35 ± 1.24	0.63	0.26 ± 0.76	0.35 ± 0.97
Izana	FTIR	8.55	1.08 ± 1.30	0.08	1.05 ± 2.26	0.37	0.73 ± 1.07	0.85 ± 1.53
Izana	O3S	4.00	**2.12 ± 1.18**	**0.00**	**1.92 ± 1.52**	**0.01**	**2.30 ± 0.87**	**1.79 ± 1.06**
Jungfraujoch	FTIR	8.56	**-1.93 ± 1.78**	**0.03**	**-1.83 ± 1.64**	**0.03**	-1.08 ± 1.34	-1.33 ± 1.86
Kiruna	FTIR	7.26	**-1.77 ± 1.48**	**0.02**	-1.26 ± 1.33	0.06	**-1.73 ± 1.15**	**-1.77 ± 1.42**
Legionowo	O3S	4.86	**-1.26 ± 1.18**	**0.04**	**-2.02 ± 1.32**	**0.00**	**-1.40 ± 1.06**	**-1.86 ± 1.05**
Lerwick	O3S	4.93	-1.01 ± 1.54	0.18	-0.77 ± 1.57	0.33	-0.96 ± 1.24	-1.02 ± 1.46

Madrid	O3S	3.96	-0.74 ± 1.24	0.25	-0.61 ± 1.29	0.38	-0.62 ± 1.22	-0.54 ± 1.39
NyAlesund	O3S	6.50	-0.75 ± 1.08	0.15	**-1.21 ± 1.14**	**0.04**	**-0.93 ± 0.91**	-0.98 ± 1.18
OHP (040)	Umkehr	11.29	0.51 ± 2.10	0.62	-0.19 ± 2.40	0.88	-0.86 ± 1.88	-0.76 ± 1.92
OHP	Lidar	6.72	**2.24 ± 1.76**	**0.01**	**3.00 ± 2.13**	**0.00**	1.90 ± 2.04	**2.48 ± 1.97**
OHP	O3S	3.85	**1.37 ± 1.26**	**0.03**	1.04 ± 1.39	0.12	**1.96 ± 1.05**	**1.67 ± 1.10**
Payerne	O3S	12.75	**-1.29 ± 1.02**	**0.01**	**-1.98 ± 1.57**	**0.01**	**-1.63 ± 0.94**	**-2.53 ± 1.22**
Sodankyla	O3S	4.17	**-1.74 ± 1.40**	**0.01**	**-1.89 ± 1.42**	**0.01**	**-1.75 ± 1.08**	**-2.02 ± 1.20**
Uccle	O3S	11.80	**1.23 ± 1.10**	**0.03**	0.83 ± 1.51	0.27	0.57 ± 0.97	0.51 ± 1.31
Valentia	O3S	4.06	1.37 ± 2.04	0.18	1.63 ± 2.68	0.23	-0.36 ± 2.41	-0.02 ± 3.20
Zugspitze	FTIR	9.68	-1.15 ± 1.82	0.22	-1.75 ± 2.51	0.16	-0.32 ± 0.60	-1.69 ± 2.40
Northern Hemisphere (80E-180E) TrOC (surface to 300hPa) Trends								
Kuala Lumpur	O3S	2.07	**2.61 ± 1.74**	**0.00**	**2.26 ± 1.91**	**0.02**	**1.86 ± 1.56**	**1.92 ± 1.55**
Rikubetsu	FTIR	3.73	-0.12 ± 1.24	0.85	-0.29 ± 2.03	0.78	-0.58 ± 1.37	-1.37 ± 1.72
Southern Hemisphere TrOC (surface to 300hPa) Trends								
Arrival Heights	FTIR	6.24	**-1.25 ± 1.20**	**0.04**		**0.00**	**-1.69 ± 1.32**	**-2.20 ± 1.48**
Fiji	O3S	2.58	-1.04 ± 1.80	0.29	0.0 ± 1.70	1.00	-1.33 ± 2.28	-1.01 ± 2.27
Irene	O3S	2.48	0.48 ± 2.36	0.68	0.49 ± 2.11	0.66	-0.16 ± 2.41	0.31 ± 2.63
Lauder	O3S	3.84	0.01 ± 0.70	0.98	0.24 ± 0.75	0.52	0.13 ± 0.61	-0.16 ± 0.79
Lauder	FTIR	10.90	**1.64 ± 0.86**	**0.00**	**1.85 ± 1.10**	**0.00**	**1.67 ± 0.86**	**1.21 ± 1.08**
Lauder (256)	Umkehr	8.97	0.38 ± 1.20	0.55	0.69 ± 1.61	0.41	0.58 ± 0.86	0.35 ± 0.92
Nairobi	O3S	3.90	0.47 ± 1.56	0.54	0.69 ± 1.62	0.40	0.75 ± 1.37	0.60 ± 1.41
Natal	O3S	3.67	0.76 ± 1.22	0.21	0.90 ± 1.83	0.33	1.04 ± 1.37	1.49 ± 1.73
Reunion	O3S	3.27	1.17 ± 1.62	0.15	0.51 ± 1.65	0.54	**1.93 ± 1.27**	**1.45 ± 1.40**
Samoa	O3S	3.31	-0.49 ± 1.10	0.35	-0.97 ± 1.60	0.23	-0.52 ± 0.99	-0.52 ± 1.03
South Pole	O3S	4.89	**-0.90 ± 0.56**	**0.00**	-0.55 ± 0.76	0.15	**-1.01 ± 0.73**	-0.36 ± 0.77

The impact of the reduced sampling frequency on the calculated trends can be assessed in Table R1, where the QR and MLR L3 trends, trend uncertainties (2 sigma values) and p values (QR) from the subsampled monthly mean time series can be directly compared. However, it should be stressed that this table only contains the trend results from one possible random subsampling to 2 daily observations a month, and different results might be obtained if the exercise is repeated with other random selections. For instance, for a time series with 10 daily values for a specific month, you have 10*9/2=45 possibilities to subsample to exactly 2 daily values for that specific month only. As we require time series with at least 120 monthly values in our study, we end up with a large number of possibilities for a time series subsampled to exactly two daily values a month. Ideally, the experiment should be executed for a large number of random subsampling strategies. This concept is nicely illustrated in Fig. 6 of Chang et al. (2024), with trends calculated from the Mauna Loa Observatory (free-tropospheric) ozone data set subsampled randomly and independently over 1000 iterations to a fixed number of samples per month (ranging between 2 and 20). For instance, for this dataset, a strategy of four samples per month yielded an accurate trend only 10% of the time, while a strategy of just two samples per month yielded an acceptable rate of zero because the subsamples either severely overestimated the trend or were not able to detect the trend. Of course, statistical power is heavily affected by the absolute magnitude of the trend and sigma values, and will be therefore site-dependent. Such an analysis for all our sites clearly falls outside the scope of this manuscript.

Taking this precaution into account, we however still compare the original L3 trends values with the trends from the subsampled datasets. It should be mentioned that our single "subsampled" dataset contains fewer sites, as the IAGOS airports DAL and ATL have too many months with only one measurement a month, so that any time series sampled to exactly two daily values a month contain too much gaps for a reliable trend estimation. From Table R1, we can conclude that, for this specific subsampling strategy, the differences between the trend values estimated from both samples are not large (mean absolute trend difference of 0.46±0.37 and 0.37±0.38 ppb/dec for QR and MLR, respectively). For both methods, only at around 5 sites, there is a trend sign reversal, but the estimated trends have large uncertainties. The most striking and consistent feature of the comparison is the higher trend uncertainties (2 sigma's) for the subsampled dataset (i.e. for 46/49 sites, or 87/92%, for QR/MLR). Also the QR p values of the "subsampled" trend estimation are higher for the majority of the sites (32 sites, 60%). As a consequence, the number of sites with trends significantly different from zero (taking a p-value lower than 0.05 as criterion) decreases, although moderately, from 22/21 (40/38%) to 16/15 (30/28%) for QR/MLR when subsampling the data to exactly two monthly measurements for calculating the monthly mean. This is not unexpected because, based on the sampling theory (Thompson, 2012), if the samples are chosen randomly and have no structural bias, the results are not expected to be biased, but smaller samples lead to larger uncertainty. As a matter of fact, the differences in trend values and trend uncertainties between the two L3 datasets are comparable with those between QR L1 and QR L3: a similar amount of sites with larger than smaller trend estimates, a mean absolute trend differences of 0.46±0.40 ppb/dec, 7 (out of 55 sites) switching trend sign, all but one sites having larger trend uncertainties (2 sigma values) for QR L3 compared to QR L1, 40 sites having larger QR L3 trend p values, the number of sites with trends significantly different from zero (taking a p-value lower than 0.05 as criterion) being 32 for QR L1 (and 22 for QR L3). On top of that, the trend (uncertainty) differences between the original and subsampled datasets lie also in the same order of magnitude than those between the two trends estimation methods (QR and MLR) used, with a mean absolute trend difference of 0.38±0.37 ppb/dec, and exactly the same number of sites having trends significantly different from zero.

We can therefore conclude that the trend uncertainty due to a monthly sampling frequency of around 2 is comparable to the trend uncertainty that is associated with the choice of the trend estimation method and with the one due to the sampling frequency (all measurements vs. monthly means) for the QR trend estimation.

We added the discussion in the previous paragraph to the supplement, below the newly added Table R1 (Table S6). In the manuscript text, the added paragraph looks like: "To further study the impact of the monthly sample numbers (SN) on the trend estimations and their uncertainties, we randomly selected for all sites two daily mean (L2) values for each month and calculated the corresponding monthly mean L3 data. Then, we estimated QR and MLR trends for both the original L3 and the "subsampled" L3 time series. As different combinations of two random samples per month are possible at the bulk of the sites, this trend sensitivity experiment should be executed for a large number of random subsampling strategies. This concept is illustrated in Fig. 6 of Chang et al. (2024), with trends calculated from the Mauna Loa Observatory (free-tropospheric) ozone data set, subsampled randomly and independently over 1000 iterations to a fixed number of samples per month (ranging between 2 and 20). Such an analysis for all our sites clearly falls outside the scope of this manuscript, and we consider only one subsampled L3 time series. The differences in the trends and their uncertainties with the full L3 time series are presented and shortly described in Fig. S6 and Table S6 of the supplementary material. In general, the mean absolute trend differences are rather modest (of the order of 0.4-0.5 ppb/dec for both QR and MLR). The most consistent feature of the comparison is the higher trend uncertainties (standard deviations and p-values) for the large majority of the sites in case of the subsampled datasets. As a matter of fact, we found that the differences in trend values and trend uncertainties between the two L3 datasets are comparable with those between QR L3 and MLR L3 and between QR L1 and QR L3 for the complete, original time series (see Table S6 and details in supplementary material). We can therefore conclude that the trend uncertainty due to a hypothetical monthly sampling frequency of 2 is comparable to the trend uncertainties associated with the choice of (i) the trend estimation method and (ii) the temporal sampling (all measurements vs. monthly means) for the QR trend estimation".

Reference: Thompson, Steven K. (2012) Sampling, 3rd Edition. John Wiley and Sons. ISBN-13: 978-0470402313

Page 8: The IAGOS profiles are integrated so that the concentration is also reported in DU. Why is this not also done for ozonesondes?

This is also done for ozonesondes. All ozonesonde partial columns are obtained by integrating the profiles, to obtain those in DU and ppb. But in case of the column-averaged tropospheric ozone mixing ratio X_{O3}, the partial ozone column amount is divided by the extent of the column. This is not done for the partial column ozone amounts in DU. All those metrics, in the two different units (DU and ppb), can be obtained at the HEGIFTOM ftp-server (details can be found on the HEGIFTOM website). The partial tropospheric ozone columns in DU have also been compared at collocated or nearby sites between different techniques (see section 4.1.1), and the results have been included in Tables S4 and S5 in the supplementary material (Tables S2 and S3 being the results for the column-averaged tropospheric ozone mixing ratio X_{O3}). In the text, we now explicitly mention that all partial ozone columns are also available in DU for ozonesondes: "To calculate tropospheric ozone columns in DU and ppb, the different ozone concentrations in the respective units at the pressure levels within a tropospheric column are integrated, and only for the case of retrieving the column-averaged tropospheric ozone mixing ratio X_{O3} divided by the extent of the column."

Section 2.5: While the inclusion of Lidar data is desirable, as it summarizes nighttime trends, I am not sure it is appropriate to compare those nighttime trends to other instruments that measure during daylight hours. Nighttime ozone trends are known to be different from daytime (e.g., Yan et al., 2018). Perhaps the authors could discuss those nighttime trends as a separate section without comparison to daytime observations. Does the fact that these are nighttime measurements or some other aspect of Lidar sensing (e.g., sensitivity to the lower atmosphere, the filling in of missing data using models) explain the biases identified in Section 4.1.1 when Lidars are compared to the other instruments?

This comment includes two very good points. First of all, the possible difference between nighttime and daytime tropospheric ozone trends. The time of day differences discussed in Yan et al., 2018 refer to surface ozone trends. Based on the frequent IAGOS FRA profiles, Petetin et al. (2016) found statistically significant diurnal variations in the mean ozone mixing ratios regardless of pressure level, although they quickly decrease with altitude (and hardly discernible above 750 hPa). Inspired by this study, we split the IAGOS FRA time series into daytime and nighttime measurements, see Fig. R2, with the start and end UTC hour of the day varying for each month separately. For your information, this is the only dataset that has a large enough sample of both the daytime and nighttime measurements for reliable trend estimation (the other IAGOS times series ATL and DAL do not fulfil this criterion). A comparison of all, daytime, nighttime QR L1 trends from 2000 for 3 different partial ozone columns (TrOC, FTOC, LTOC) are shown in the Table R2 here below. It should be noted that there is a large difference between the daytime and nighttime partial tropospheric ozone trends: the daytime trends are close to zero (slightly positive), while the nighttime trends are strongly positive (between 1.61 and 2.04 ppb/dec, depending on the partial ozone column). The largest difference between the daytime and nighttime trends occur for the LTOC (from the surface to 700 hPa), which aligns with the larger diurnal cycle found at these levels by Petetin et al. (2016). Apparently, those large LTOC trend differences between day and night also contribute to the trends differences for the entire TrOC at Frankfurt airport. However, the vertical tropospheric ozone trends for 1994-2019 have been calculated and compared at different pressure levels between the entire and daytime IAGOS FRA observations in the supplement (Fig. S8) of Chang et al. (2022), and no clear trends differences arise between the entire and daytime sample. Therefore, much more analysis on the spatial and temporal sampling differences between daytime and nighttime IAGOS FRA observations is needed to understand the partial ozone column trend differences between those two subsets, before looking at differences in the chemistry producing or destroying ozone during day or night. Ideally, this would be the subject of a separate paper and lies beyond the scope of this manuscript.

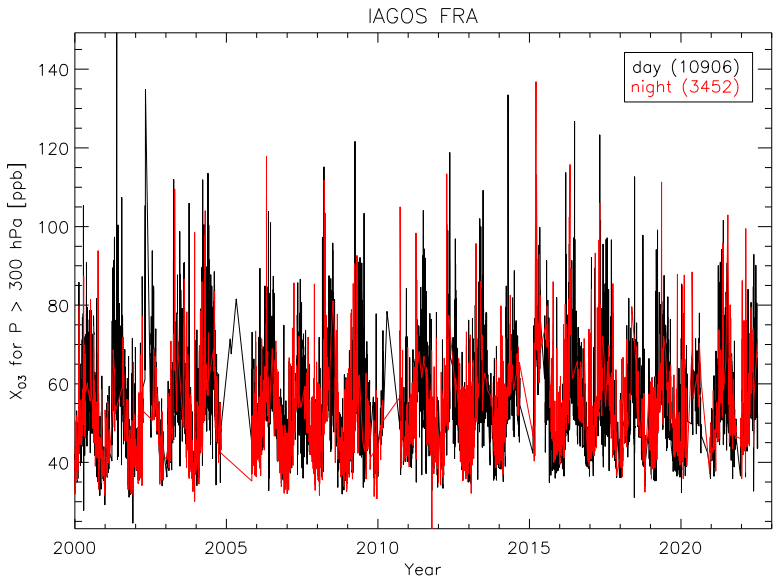

Fig. R2: IAGOS FRA TrOC all measurements (L1) time series, split between daytime observations (black) and nighttime observations (red). The begin and end UTC hour of the day have been varied from month to month. The numbers between brackets in the legend denote the sample numbers for the entire time series (since 1994), and missing values for the TrOC have been filtered out.

Table R2: QR L1 TrOC, FTOC, and LTOC 2000-2022 trend estimates, with uncertainties, calculated for the entire, daytime and nighttime time series at IAGOS Frankfurt. The first columns mention the different sample numbers, and the mean TrOC values (with standard deviation) for these different datasets.

			TrOC			FTOC			LTOC		
	N	Mean (ppb)	Trend (ppb/dec)	2 σ	p value	Trend (ppb/dec)	2 σ	p value	Trend (ppb/dec)	2 σ	p value
all	14358	54.12±11.36	0.65	0.36	0.00	0.59	0.47	0.01	0.57	0.41	0.01
day	10906	55.31±11.26	0.28	0.40	0.16	0.22	0.49	0.38	0.12	0.47	0.60
night	3452	50.37±10.84	1.84	0.57	0.00	1.61	0.70	0.00	2.04	0.65	0.00

Second, you mention some other aspects of Lidar sensing that might explain differences in mean values and trends of the lidar datasets with respect to other techniques. We should note that both TMF and OHP lidars have their lowest data points at around 3 km above the surface, often even higher. The Lidars at these sites measure basically only the free tropospheric column, so the best lidar metric is the FTOC metric (from 700 to 300 hPa). When we compare those mean FTOC differences, we found that the TMF lidar measurements reveal a 1-10% positive FTOC difference with IAGOS, and the OHP lidar a 5% positive FTOC difference with ozonesondes, independent of the chosen unit (mixing ratio or DU). Those numbers are lower and more consistent for the different units when compared to the TrOC comparisons, where the differences are positive in terms of ozone mixing ratio, but negative in terms of DU. This points to an origin for the lidar differences that is related to the conversion of units, as seem the case for the FTIR biases. Both techniques require meteorological auxiliary data from reanalyses to convert from DU to column-averaged ozone mixing ratios in ppb.

We included this discussion at several locations in the manuscript. First, in section 4.1.1, we write “The TMF lidar measurements reveal a positive TrOC difference with IAGOS and the OHP lidar has a positive TrOC difference with Umkehr and ozonesondes, see Table S3. We should however note that both those lidars have their lowest data points at around 3 km above the surface, so the best lidar partial ozone column metric for comparison with other techniques

is the free-tropospheric ozone column (FTOC) between 700 to 300 hPa. With this metric, also positive FTOC differences with IAGOS (TMF) and ozonesondes (OHP) are found."

In the same section, we add "In contrast to most of the co-located techniques, lidar ozone measurements are nighttime measurements. Based on the frequent IAGOS FRA profiles, Petetin et al. (2016) found statistically significant diurnal variations in the mean ozone mixing ratios regardless of pressure level, although they quickly decrease with altitude (and hardly discernible above 750 hPa). Therefore, differences in daytime and nighttime mean ozone mixing ratios might partially contribute to the TrOC and FTOC differences between lidar and other co-located techniques."

The discussion of the trend differences between different techniques at collocated sites (section 4.3.1) has been extended with the following paragraph: "At OHP, the lidar nighttime TrOC trend estimate is very close to the ozonesonde daytime TrOC trend, but those positive trends differ largely from the negative Umkehr daytime trend value. Therefore, at first sight, the impact of sampling during day or night on the trend estimations seems rather limited. However, if we estimate the daytime (76% of the observations) and nighttime partial ozone column trend estimates from the IAGOS FRA time series, the close to zero daytime trends are substantially different from the large positive nighttime trends (between 1.61 and 2.04 ppb/dec, with the largest values for the lower tropospheric ozone column trends). This finding requires further investigation, e.g. to check the extent of a possible sampling bias (temporal or spatial) between both subsets on the trend estimations."

Figures:

Figure 6. The legend in the upper left corner is difficult to use. A few suggestions: 1) box off the legend so it does not appear to be another data point, and 2) provide more than one length/concentration for an easier visual reference.

Thank you for these very good suggestions. These have been implemented in the updated figure 6.

References:

Chang, K.-L., Cooper, O. R., Gaudel, A., Petropavlovskikh, I., and Thouret, V.: Statistical regularization for trend detection: an integrated approach for detecting long-term trends from sparse tropospheric ozone profiles, Atmos. Chem. Phys., 20, 9915–9938, https://doi.org/10.5194/acp-20-9915-2020, 2020.

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?, Atmos. Chem. Phys., 22, 14751–14782, https://doi.org/10.5194/acp-22-14751-2022, 2022.

Lu, X., Zhang, L., Zhao, Y., Jacob, D. J., Hu, Y., Hu, L., Gao, M., Liu, X., Petropavlovskikh, I., McClure-Begley, A., and Querel, R.: Surface and tropospheric ozone trends in the Southern Hemisphere since 1990: possible linkages to poleward expansion of the Hadley circulation, Sci. Bull., 64, 400–409, https://doi.org/10.1016/j.scib.2018.12.021, 2019.

Wang, H., Lu, X., Jacob, D. J., Cooper, O. R., Chang, K.-L., Li, K., Gao, M., Liu, Y., Sheng, B., Wu, K., Wu, T., Zhang, J., Sauvage, B., Nédélec, P., Blot, R., and Fan, S.: 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, Atmos. Chem. Phys., 22, 13753–13782, https://doi.org/10.5194/acp-22-13753-2022, 2022.

Yan, Y., Lin, J., and He, C.: Ozone trends over the United States at different times of day, Atmos. Chem. Phys., 18, 1185–1202, https://doi.org/10.5194/acp-18-1185-2018, 2018.

Thank you. Most of these were already in the manuscript.