Ozone stratospheric trends from regional Bayesian composite of ground-based partial columns

Louis Mirallie[1, 2, 3], Eliane Maillard Barras[1], Caroline Jonas [4], Corinne Vigouroux [4], Roeland Van Malderen [5], Irina Petropavlovskikh[6, 7], Sophie Godin-Beekmann S.[8], Thierry Leblanc[9], Wolfgang Steinbrecht[10], Antoine Vadès[11], Rolf Ruefenacht[1], Alexander Haefele[1], Gunter Stober[2, 3], Peter Effertz [6, 7], Julian Gröbner[12], Gerard Ancellet [8], María Cazorla[13], Petra Duff[14], Matthias M. Frey[15], Michael Gill[16], James W. Hannigan[17], Nicholas Jones[18], Rigel Kivi[19], Raphael Köhler[20], Bogumil Kois[21], Debra E. Kollonige[22, 23], Emmanuel Mahieu[24], Glen McConville[7], Johan Mellqvist[25], Gary Morris[7], Isao Murata[26], Tomoo Nagahama[27], Gerald E. Nedoluha[28], Shin-Ya Ogino[29], Richard Querel[30], Ryan M. Stauffer[22], Wolfgang Stremme[31], Kimberly Strong[14], Ralf Sussmann[32], Anne M. Thompson[22, 33], Yana Virolainen[34]

[1] Federal Office of Meteorology and Climatology, MeteoSwiss, Payerne, Switzerland
[2] Institute of Applied Physics, Microwave, University of Bern, Bern, Switzerland
[3] Oeschger Center for Climate Change Research, Microwave Physics, University of Bern, Bern, Switzerland
[4] Royal Belgian Institute for Space Aeronomy (BIRA-IASB), Uccle, Belgium
[5] Royal Meteorological Institute, Uccle, Belgium
[6] Cooperative Institute for Research in Environmental Sciences, University of Colorado, Boulder, CO, USA
[7] NOAA Global Monitoring Laboratory, Boulder, CO, USA
[8] LATMOS, Sorbonne Université, UVSQ, CNRS, Paris, France
[9] Jet Propulsion Laboratory, California Institute of Technology, Wrightwood, California, USA
[10] Deutsche Wetterdienst, Hohenpeißenberg, Germany
[11] Formerly at Federal Office of Meteorology and Climatology, MeteoSwiss, Payerne, Switzerland
[12] Physikalisch-Meteorologisches Observatorium Davos World Radiation Centre, Dorfstrasse 33, 7260 Davos Dorf, Switzerland
[13] Instituto de Investigaciones Atmosféricas, Universidad San Francisco de Quito USFQ, Quito, Ecuador
[14] Department of Physics, University of Toronto, Toronto, ON, Canada
[15] Institute for Meteorology and Climate Research Atmospheric Trace Gases and Remote Sensing (IMKASF), Karlsruhe Institute of Technology (KIT), Karlsruhe, Germany
[16] Met Éireann Valentia Observatory, Cahersiveen, Co. Kerry, Ireland
[17] National Center for Atmospheric Research, Boulder, CO, USA
[18] University of Wollongong, Wollongong NSW, Australia
[19] Space and Earth Observation Centre, Finnish Meteorological Institute, Sodankylä, Finland
[20] Alfred Wegener Institute, Helmholtz Centre for Polar and Marine Research, Potsdam, Germany
[21] Institute of Meteorology and Water Management – National Research Institute (IMGW-PIB), Poland
[22] Earth Sciences Division, NASA Goddard Space Flight Center, Greenbelt, Maryland, USA
[23] ADNET Systems, Inc., Bethesda, Maryland, USA
[24] Department of Astrophysics, Geophysics and Oceanography, UR SPHERES, University of Liège, Liège, Belgium
[25] Space Earth and Environment , Chalmers University of technology, 41296 Göteborg, Sweden
[26] Graduate School of Environmental Studies, Tohoku University, Sendai, Japan
[27] Institute for Space-Earth Environmental Research, Nagoya University, Nagoya, Japan
[28] Naval Research Laboratory, 4555 Overlook Avenue, SW, Washington, D.C. 20375, USA
[29] Japan Agency for Marine-Earth Science and Technology (JAMSTEC), Research Institute for Global Change, Yokosuka, Japan
[30] Earth Sciences New Zealand (ESNZ), Lauder, New Zealand
[31] Instituto de Ciencias de la Atmósfera y Cambio Climático, Universidad Nacional Autonoma de Mexico (UNAM), Mexico City, MEXICO
[32] Karlsruhe Institute of Technology (KIT), IMKIFU, Garmisch-Partenkirchen, Germany

[33] University of Maryland - Baltimore County, Baltimore, Maryland, USA
[34] St.Petersburg State University, St.Petersburg, Russia
Correspondence to: Louis Mirallie (louis.mirallie@meteoswiss.ch)

Abstract. Large uncertainties and variability prevent the detection of statistically significant ozone trends from individual ground-based instruments in the lower stratosphere. Available merging studies are typically performed by latitude bands on satellite-based data records. This study derives correlation-based regional composites of ground-based time series towards reducing trends uncertainties.

We address fundamental heterogeneities resulting from grouping individually homogenized ground-based datasets to enable robust merging. Uneven temporal and vertical resolutions of five ozone measurement techniques (Ozonesondes, FTIR, Dobson Umkehr, Lidar and Microwave radiometers) are handled by integrating monthly mean ozone profiles in two sets of four independent partial columns. Spatial heterogeneity is resolved by defining coherent regions using the Copernicus Atmosphere Monitoring Service (CAMS) reanalysis. Regional time series are merged by the BAyeSian Integrated and Consolidated (BASIC) algorithm, adapted to consider propagated measurement uncertainties and the agreement between individual time series by Principal Component Analysis (PCA). Trends for the 2000–2024 period are then estimated by Multiple Linear Regression using the LOTUS model.

We compare BASIC with a conventional weighted mean. While the weighted mean fails to capture variability during periods of low instrument consensus, BASIC produces a more representative time series by robustly handling outliers. Accordingly, BASIC reduces average uncertainties of the trend estimates by 15.3% relative to the weighted-mean approach. Our results confirm robust positive trends in the upper stratosphere and show predominantly negative significant regional trends in the middle and lower stratosphere. This study establishes a consolidated, global ground-based reference to be used for comparison with global satellite-based ozone trends.

1 Introduction

The Montreal Protocol and its subsequent Amendments have successfully reduced the atmospheric concentrations of ozone depleting substances, initiating a slow recovery of the ozone layer (WMO, 2022). However, this recovery is spatially heterogeneous and remains vulnerable to emerging threats such as unregulated emissions (Rigby, 2019) and climate change (Wang, 2025), emphasizing the need for continuous monitoring of both stratospheric and tropospheric ozone.

The clear recovery signal in the upper stratosphere (WMO 2022) contrasts with the evolution of ozone in the lower stratosphere, which remains uncertain (Diemüller, 2021; Szelag,2020; Millan, 2025). Statistically non-significant negative ozone trends have been estimated from measurements in the lower stratosphere while models are showing the opposite (Godin-Beekmann et al., 2022; WMO, 2022). Moreover, the large uncertainties of the trend estimates in the lower stratosphere, possibly coming from the rising tropopause in the tropics (Thompson et al., 2025), prevent definitive conclusions about ozone changes in this vertical domain. We note also that trends from individual ground-based instruments

80 can be significantly different, even if the instruments are collocated (Petropavlovskikh, 2019; Godin-Beekmann, 2022; Björklund, 2024) and that large discrepancies are also reported on trend estimates from satellite-based data records (Sofieva et al., 2025).

Nevertheless, profile trend estimation is crucial to understand the recovery processes. Ozone changes are dominated by different chemical and dynamical processes depending on the altitude (WMO, 2022). Hence, the monitoring of the vertical structure of the changes is an essential addition to that of the integral measurement, i.e. the total ozone column.

85 Several considerations aimed at reducing the uncertainties in ozone profile trends can be found in the literature. The merging of data records has been extensively used in the field of satellite-based instruments, with the merging of gridded satellite-based data records (Sofieva, 2021), the merging of homogenized satellite-based data records (Davis, 2016; Arosio et al., 2019) or the merging of debiased ozone records in Froidevaux et al (2015). In the OCTAV-UTLS APARC activity, Millan et al (2025) report the ozone profile in several coordinate systems to divide the measurements into vertical domains. This

90 allows the data in each domain to be affected by the same dynamical processes, reducing their variability, which is critical for trends detection particularly in UTLS where opposite ozone changes cancel each other in standard pressure coordinates. Integration of the vertical profiles into partial columns can also be used to partially compensate for the significant uncertainties of a profile, as it averages out its vertical variability. This method is widely used in the validation of low vertical resolution instruments with higher resolution instruments (e.g. Kramarova et al., 2013, Frith et al., 2020) or data

95 assimilation products (e.g. Innes et al. 2019, Emili et al., 2014).

Satellites provide excellent geographical coverage, although their life-time is limited (e.g. SBUV series of remote sensors, SAGEII-III-IV, ENVISAT). On the other hand, ground-based instruments are geographically rather sparse and unequally distributed, but their time series are continuous (Stübi et al., 2017) and the resolution of some of them in the lower stratosphere is sufficient: less than 1km for Lidars, a few hundred meters for ozonesondes (Smit et al., 2020; Leblanc et

100 al.,2016). With the foreseen reduction of the number of limb-viewing satellite instruments in the coming years (data desert in Salawitch et al., 2025), the study of ozone recovery will rely more heavily on ground-based measurements to help fill the data gaps that will emerge for ozone and other essential climate variables.

There is therefore a sustained need for ground-based global coverage of ozone measurements. As individual ground-based stations provide limited spatial coverage, a merging of ground-based datasets representative of similar ozone variability is

105 sometimes necessary to increase the effective coverage of the merged time series. Furthermore, by combining all existing information into a composite, the uncertainty of the resulting time series, and thus of its trend estimate, can likely be reduced (e.g. Ball et al., 2017; Arosio et al., 2019; Sofieva et al., 2021; Keppens et al, 2025).

In this study we integrate heterogeneous ground-based ozone datasets into common partial columns, and define spatially

110 coherent regions with correlations derived from a representativeness analysis. We merge time series into a composite using a Bayesian algorithm (BASIC, Ball et al., 2017), and estimate 2000–2024 trends with the MLR model from LOTUS (Petropavlovskikh et al., 2019; Godin-Beekman et al., 2022) defined within the LOTUS/APARC activity (Long-term Ozone

Trends and Uncertainties in the Stratosphere) with consideration of the propagated monthly measurement uncertainties. We compare the Bayesian approach to conventional weighted mean, and provide regional partial-column trend estimates.

115 The paper is organized as follows: ground-based instrument datasets, partial column definitions and determination of uncertainties are described in Section 2. Section 3 is dedicated to the methodology and describes the determination of the regions with the representativeness study based on the CAMS reanalysis, the Bayesian merging method and the MLR trend estimation. Results are detailed in Section 4, where regional composites are computed and their 2000-2004 trends estimated for regions covering a part of the globe. Finally, conclusions are drawn in Section 5.

120

2 Data Records

The data records used in this study are mostly part of the Network for the Detection of Atmospheric Composition Change
125 (NDACC). The NDACC ozone profiles records are made of five instrument types: Ozonesondes (O3S), Fourier-Transform InfraRed spectrometers (FTIR), Lidars, Microwave Radiometers (MWR) and Dobson (Umkehr) spectrophotometers. A map of all available measurement stations by instrument type is illustrated in Figure 1, and a table describing all the instruments names, coordinates and data availability is in supplement 1.

Figure 1: Map of all available measurement stations by instrument type: ozonesonde in red, LIDAR in grey, MWR in black, FTIR in cyan and Dobson Umkehr in green.

We use harmonized data records and homogenize the propagation of measurement errors. The homogenization of the time series ensures the artifacts, jumps and drifts have been carefully considered and corrected where necessary. This harmonization work was performed within the APARC/LOTUS and TOAR-II/HEGIFTOM (Harmonization and Evaluation of ground-based Instruments for Free Tropospheric Ozone Measurements) initiatives. As a result, we assume that differences between the homogenized time series are limited to measurement (random) errors and offsets caused by variations in spatial representativeness (see section 3.1).

The instruments differ significantly in their measurement principles, uncertainties and vertical resolutions. For a consistent merging of the datasets, a common basis is required, on which instruments datasets are comparable. This is achieved by integrating the profiles over defined Partial Columns (PC), which will be described in Section 2.1. The uneven resolutions are addressed by ensuring that at least one degree of freedom per instrument is available in each defined partial column. Figure 2 illustrates the vertical resolution of the instruments as the number of horizontal bands within their vertical ranges, alongside the two sets of partial columns defined for this study (see Section 2.1).

The instruments also differ in temporal resolution as follows: Ozonesondes are launched typically 2 to 12 times per month, Dobson Umkehr provides bi-daily measurements at most, Lidar, FTIR and MWR can provide several profiles per hour, nighttime for the Lidar, daytime for the FTIR and both daytime and nighttime for the MWR. The diurnal cycle is not expected to contribute given its limited amplitude, maximum in the upper stratosphere, reported to be generally below 4% (Sauvageat et al., 2023). To resolve these disparities, we aggregate all data of each single data record into monthly means (L3). Furthermore, to ensure consistency, all monthly mean partial columns are expressed in Dobson Units (DU).

150

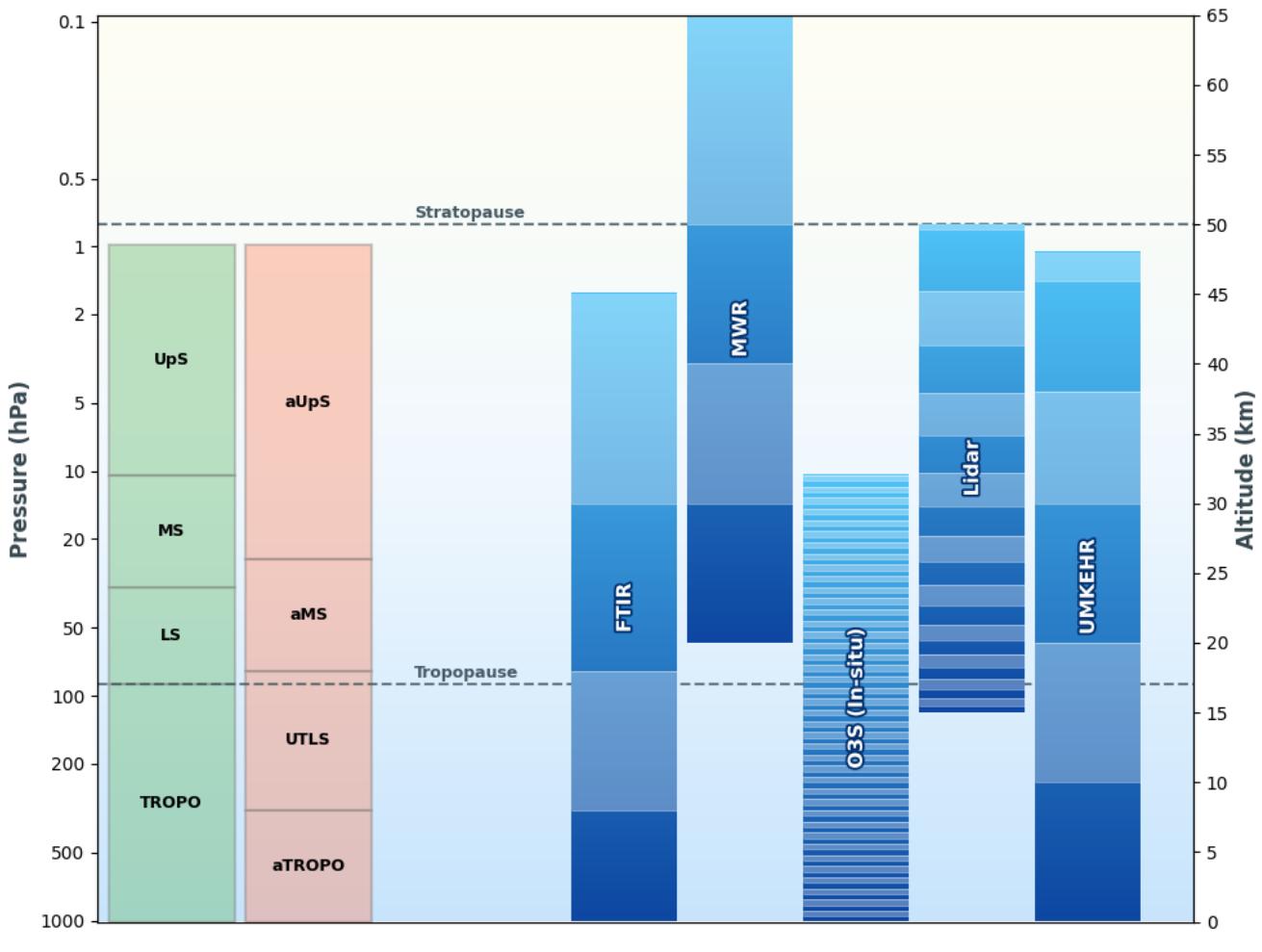



Figure 2: Representation of the two sets of partial columns used on the left-hand side and of the instruments approximate vertical resolutions and ranges on the right-hand side (adapted for ground-based ozone measurements from NDACC (link in table S1). Pressure ranges values are given in Table 1.

Finally, a consistent methodology is applied to propagate uncertainties from individual measurements (L1) to the monthly
155   means (L3) used for trend analysis. The uncertainty of the daily mean (L2) is calculated as the L1 measurement uncertainty divided by the square root of the number of measurements within that day. Then, the monthly mean uncertainty (L3) is calculated by dividing the daily L2 uncertainty by the square root of the number of measurement days in the month. Specific L1 uncertainty values are detailed in the following subsections.

2.1 Partial Columns

160   Small-scale and short-term variabilities are poorly characterized in regression analyses. This leads to large uncertainties in trend estimates when derived from highly resolved vertical profiles. Vertical integration may counteract this effect by

averaging out some variability, if the partial columns are selected for that purpose. However, fine vertical evolution can also be averaged out by vertical integration. Focusing on reducing the uncertainties of the trends estimates, we have chosen to estimate the trends of ozone on two sets of four partial columns defined as follows (see Table 1):

165
- the standard set ("oPC"), which can be found in Ball et al. (2018 and 2020) with a distinction for stations with latitudes inside and outside the 30° band.
- the alternative set ("aPC"), which has a pure troposphere and a UTLS, in order to force the variability around the tropopause into a single partial column.

Partial Columns (oPC) :

Upper Stratosphere (UpS)	10 - 1 hPa	32 km - 48 km
Middle Stratosphere (MS)	32 - 10 hPa	24 km - 32 km
Lower Stratosphere (LS)	100*/147 - 32 hPa	17*/13 km - 24 km
Troposphere (TROPO)	surface - 100*/147 hPa	0 - 17*/13 km

Alternative Partial Columns (aPC)

Upper Stratosphere (aUpS)	20 - 1 hPa	26 km - 48 km
Middle Stratosphere (aMS)	64 - 20 hPa	18 km - 26 km
UTLS	150/200/300/400** - 32 hPa	12/11/9/8** km - 18 km
Troposphere (aTROPO)	surface - 150/200/300/400** hPa	0 - 12/11/9/8** km

170 Table 1: oPC and aPC are two sets of partial columns defined on pressure levels (approx. heights in km) with latitude distinction. The two sets of partial columns can be seen on the left-hand side of Figure 1. * Latitude < 30°, ** Latitude < 15/30/60/90°

For all instrument types, the ozone amount in each partial column has been computed as the sum of ozone content in function of the pressure level within the partial columns' vertical ranges. Ozonesondes do not provide any values in the upper stratosphere, MWR do not provide any tropospheric nor UTLS ozone values and Lidars provide a complete stratosphere as described in section 2.2.3. Monthly means are computed from daily mean values. The minimum number of measurement days per month was set to three.

2.2 Instruments

2.2.1 Ozonesondes

Ozonesondes, launched with weather balloons, are small, light-weight instruments that measure the vertical ozone profile up to about 30-35 km altitude (10-5 hPa), based on the titration of ozone in a neutral buffered potassium iodide sensing solution. This measurement technique has a stated precision better than $\pm$ (3 %–5 %) and an accuracy of about $\pm$ (5 %–10 %) for up to 30 km altitude (Smit et al., 2021, 2024). As major contributors to uncertainties in ozone trends are discontinuities and biases in the long-term records of ozonesonde sites due to e.g. changes in ozonesonde type, sensing solution, and preparation/processing, the ozonesonde data has been homogenized as described in Smit et al. (2021), Van Malderen et al. (2025a), and references therein. This homogenization (normally) also involves the provision of an uncertainty estimate for every single measurement.

To calculate partial ozone columns from the ozonesonde data, the profiles are simply integrated between the upper and lower boundaries of the defined atmospheric layers, if the percentage of missing (intermediate) profile data in the partial column is not higher than 20%. The uncertainties of the partial ozone columns are obtained by summing up the individual uncertainties of the ozone concentration measurements. Most sites typically launch once a week, but the average monthly launch frequencies vary roughly within the ozonesonde network between 2 to 12. The uncertainties on the monthly means have been computed as described above (see section 2) and are around 5 to 6% on average.

2.2.2 FTIR

The FTIR (Fourier Transform InfraRed) ozone measurements are uniformly performed in NDACC with Bruker high-resolution spectrometers (except for Toronto, which uses an ABB Bomem instrument). This remote sensing technique retrieves trace gas concentrations from solar absorption spectra, requiring daylight and cloud-free conditions. The FTIR ozone retrievals have been standardized within the IRWG (InfraRed Working Group, https://www2.acom.ucar.edu/irwg). Details on the instruments, retrieval codes and principles can be found in Vigouroux et al. (2015). In summary, ozone total columns can be retrieved from the area of ozone absorption lines by fitting spectral regions modelized using a spectroscopic database (HITRAN), climatological a priori profile information on all absorbing gases, and a radiative transfer model. Profiles of limited vertical resolution can be obtained from the line shapes (pressure and temperature dependent) by using regularization techniques. The vertical resolution and degrees of freedom for signal (DOFS) are described by the averaging

205   kernels. For ozone, the sensitivity covers the surface up to about 48 km, with about 4 DOFS: one in the troposphere and three in the stratosphere (Vigouroux et al., 2015).

Except for Altzomoni and Izana, the present paper uses the latest improved version of NDACC FTIR ozone data (Bjorklünd et al., 2024), in which the spectral windows have been optimized to avoid water vapor line interferences while the spectroscopic parameters have been updated to HITRAN 2020 (Gordon et al., 2022)

210   The uncertainty characterization is based on optimal estimation (Rodgers, 2000; and Vigouroux et al. 2015 for details on its application to FTIR ozone retrievals). The random part is dominated by the measurement noise for total columns (about 1.0-1.4% for an individual measurement) and by the smoothing random uncertainty for partial columns (about 4-6%). The systematic uncertainty is dominated by spectroscopy, especially for total columns (2-3%), and by the instrumental line shape and the temperature a priori profile errors for the partial columns (5-7%).

215   **2.2.3 Lidar**

Lidar (Light Detection And Ranging) is a remote sensing technique based on the interaction of coherent light source with the atmosphere. The use of pulsed laser sources enables range-resolved measurements (active remote sensing). Ozone measurements by lidar are performed using the DIAL (Differential Absorption Lidar) technique, which uses the simultaneous emission of two laser wavelengths absorbed differently by ozone in the atmosphere. For stratospheric ozone,

220   the weakly-absorbed wavelength of 308 nm and non-absorbed wavelength of 355 nm are chosen to optimize sensitivity through the entire stratospheric ozone layer (10-50 km). The selection of this wavelength pair is linked to the simultaneous decrease in the upper stratosphere of the ozone number density and of the atmospheric number density that provides the backscatter radiation (Godin-Beekmann et al., 2003; Leblanc et al., 2016b). After a few typical corrections (e.g., non-linearity, background noise) of the raw lidar signals, ozone number density is retrieved by differentiating with respect to

225   altitude the logarithm of the ratio of the corrected signals at the absorbed and non-absorbed wavelength. To account for the rapid decrease of the signal-to-noise ratio in the high-altitude range, a low-pass filter is used, which decreases the effective vertical resolution of the ozone profile (Leblanc et al., 2016a). Depending on the laser power and the altitude of the station, ozone vertical distribution is generally retrieved from 10 to 45-50 km altitude, with an effective vertical resolution ranging from less than 1 km to more than 5 km, and with total uncertainty varying from a few percent in the lower stratosphere to

230   more than 15% at 50 km, depending on the power of the emitted laser radiation and choice of vertical resolution (Godin-Beekmann et al., 2003; Leblanc et al., 2016b). Stratospheric ozone measurements are performed during the night to avoid high background noise from the solar radiation, and in quasi cloud-free conditions to avoid interference and absorption by clouds. Only the presence of cirrus clouds is tolerated for the measurements. The lidar instruments used in this study have been running in routine mode 2 to 8 hours per night, 1 to 5 nights per week over a time span of several decades.

2.2.4 Dobson Umkehr

The Umkehr is an observational method designed to collect zenith sky observations at two wavelengths in the ultraviolet part of the spectrum at varying solar zenith angles during sunrise or sunset. An optical estimation technique is used for retrieving the vertical profile of ozone from the observations collected by Dobson spectrophotometer instruments at wavelengths centered at 311.5 and 332.4 nm (i.e. Petropavlovskikh et al, 2005; Petropavlovskikh et al., 2008; Petropavlovskikh et al. 2009; Petropavlovskikh et al., 2011). The Brewer ozone profiles are retrieved using similar techniques, with zenith sky observations at ~310 and ~326 nm (Petropavlovskikh et al. 2011). Dobson Umkehr data have been homogenized by assessing instrumental artifacts that created step changes in the records (Petropavlovskikh et al. 2022; Maillard Barras et al. 2022). The total uncertainty characterization of Umkehr profile retrievals is based on the Rodgers (2000) optimal estimation technique and includes measurement and vertical smoothing errors. The Dobson and Umkehr retrievals are optimized for monthly mean ozone values. Therefore, for estimation of the error of a single profile retrieval, we make assumptions based on the monthly mean (MM) errors and the fact that MM (or L3) errors are daily profile errors over the square root of the number of days in a month. Uncertainties of the Umkehr ozone profile retrievals are provided by a monthly covariance matrix. To calculate uncertainties of four large partial columns for this paper, a root sum of squares method is used to combine uncertainties of each Umkehr partial column that has its pressure limits within the pressure boundaries of larger partial columns.

2.2.5 Microwave radiometer (MWR)

Microwave radiometry measures the intensity spectrum of the ozone emission line at 142.175 GHz or 110.836 GHz to retrieve ozone profiles in the stratosphere and the lower mesosphere (20 to 65 km) by the optimal estimation method (Rodgers, 2000). The MWR is a ground-based all-weather instrument. The spectral distribution is measured by a fast-Fourier-transform spectrometer or by a filter-bank spectrometer depending on the instrument. The ozone altitude distribution is retrieved from the pressure-broadened line shape. Ozone values are reported as a mixing ratio (ppmv) as a function of pressure and converted to Dobson Units (DU) for this study.

The total uncertainty is calculated for each retrieved profile accounting for measurement noise, tropospheric attenuation, calibration load temperatures, spectroscopy, atmospheric temperature profile and smoothing as sources of uncertainty. Both observation and smoothing errors are considered random uncertainties. The total uncertainty is dominated by systematic errors below 2 hPa and by the measurement noise above.

The total uncertainty is 9 %–15 % for the Bern MWR, 7 %–12 % for the Payerne MWR, 5%- 9% for the Lauder MWR and 7%–9% for the Mauna Loa MWR.

The measurements are nearly continuous, and the vertical resolution ranges from 9 km in the lower stratosphere to 15 km in the low mesosphere. Details on the retrieval process and the uncertainty budget can be found in Sauvageat et al. (2022).

The Bern and the Payerne MWRs have been harmonized in 2022. Both data records have been homogenized in 2010 for an upgrade to an AC240 FFTS spectrometer (Maillard Barras et al., 2020). The Mauna Loa data record uses a similar retrieval routine and has been homogenized to account for the low bias caused by the use of an AC240 spectrometer during the period in 2015-2017 (Sauvageat et al, 2021). The Lauder MWR data record ended in 2016 and should resume in March 2026. The four MWR data records are included in the Network for the Detection of Atmospheric Composition Change (NDACC).

3 Methods

3.1 Definition of regions

Regional groups used for the BASIC's merging are determined following a representativeness study based on CAMS global reanalysis (EAC4) monthly averaged fields of ozone (Innes et al, 2019). This type of study was proposed in Weatherhead et al. (2017), where a satellite dataset was used to evaluate the representativeness - or spatial coverage - of a ground-based network measuring the UTLS temperature. In the context of tropospheric ozone, it was further considered in Van Malderen et al (2025b) based on the CAMS data, and similarly applied to total, tropospheric and stratospheric ozone in Jonas et al. (2026). Here we use spatial correlations of the CAMS partial columns of ozone at each location of ground-based sites to determine if records from geographically selected sites can be merged together to determine representative trends for that region. As we find, the size of these regional groups highly depends on the altitude of the partial columns. Because the two sets of partial columns largely overlap in altitude range, the representativeness of one partial column is valid for the other one, so we only run this representativeness study on the original oPC definition and not on the alternative set aPC. To avoid having the correlations dominated by the seasonal cycle, we computed them using the CAMS anomalies time series of PC ozone monthly means. Since the CAMS data is only available from 2003 to 2024, it was the time range used to compute the correlations. In this study we have 64 distinct ground-based sites (given the grid precision of CAMS, i.e. of 0.75∘ lat. × 0.75∘ lon., we cannot distinguish between sites too close to each other, 0.75° varies between 82.9 to 83.9 km depending on the latitude). For the four partial columns, we calculate the spatial correlation of monthly anomalies of CAMS between each pair of sites. We obtain correlation tables, from which we construct regional groups such that all sites within a group correlate with a Pearson coefficient r larger than 0.75. According to Weatherhead et al. (2017), correlations of r>0.7 are referred to as "well correlated" and r>0.9 and above as "strongly correlated".

The procedure of defining the regional groups is presented in detail here for the example of the continental Europe ground-based sites in the lower stratospheric partial column. The corresponding correlation table is given in Figure 3. This table leads to the North Sea group (Valentia, Uccle, De Bilt, Lerwick), the Central Europe group (Payerne, Bern, Zugspitze, Arosa/Davos, Hohenpeissenberg, Uccle, Bremen, OHP, De Bilt, Jungfraujoch, Legionowo) and the South Europe group (Madrid, OHP, l'Aquila). By plotting the correlation maps with respect to one specific site, we can visualize the spatial extent of these groups, as in Figure 4 and 5 for Zugspitze FTIR in the lower and upper stratosphere respectively. The group is defined as the highest number of instruments that are all two-by-two correlated with the threshold defined above. In the

case of the group central Europe LS, it could not be fused with south Europe because Madrid and L'Aquila are poorly correlated to Bremen, De Bilt and Valentia.

300 Such an analysis is performed for all ground-based sites around the globe and for the tropospheric, lower, mid and upper stratospheric partial columns. For each partial column, we do not consider groups that contain less than 3 instruments measuring in that partial column range because in that case, the BASIC merging procedure described in section 5 does not bring any added value compared to a simple weighted mean of the time series.

	Bremen	Legionowo	De Bilt	Valentia	Uccle	Hohenpeiss.	Zugspitze	Payerne	Arosa/Davos	Jungfraujoch	OHP	I Aquila
Legionowo	0.893											
De Bilt	0.986	0.83										
Valentia	0.754	0.527	0.821									
Uccle	0.964	0.806	0.991	0.818								
Hohenpeissenberg	0.895	0.863	0.894	0.622	0.922							
Zugspitze	0.895	0.863	0.894	0.622	0.922	1						
Payerne	0.872	0.789	0.895	0.676	0.936	0.981	0.981					
Arosa/Davos	0.874	0.826	0.884	0.633	0.921	0.995	0.995	0.994				
Jungfraujoch	0.874	0.802	0.892	0.661	0.932	0.987	0.987	0.999	0.997			
OHP	0.743	0.677	0.775	0.602	0.833	0.914	0.914	0.957	0.945	0.954		
I Aquila	0.672	0.713	0.674	0.468	0.724	0.882	0.882	0.882	0.899	0.888	0.93	
Madrid	0.507	0.419	0.561	0.551	0.63	0.684	0.684	0.765	0.731	0.754	0.877	0.772

305

Figure 3: Correlations in the lower stratospheric partial column for Continental Europe sites.The cell colors vary from yellow to blue based on the correlation. The correlation (r) above (resp. below) 0.75 is written in black (resp. white).

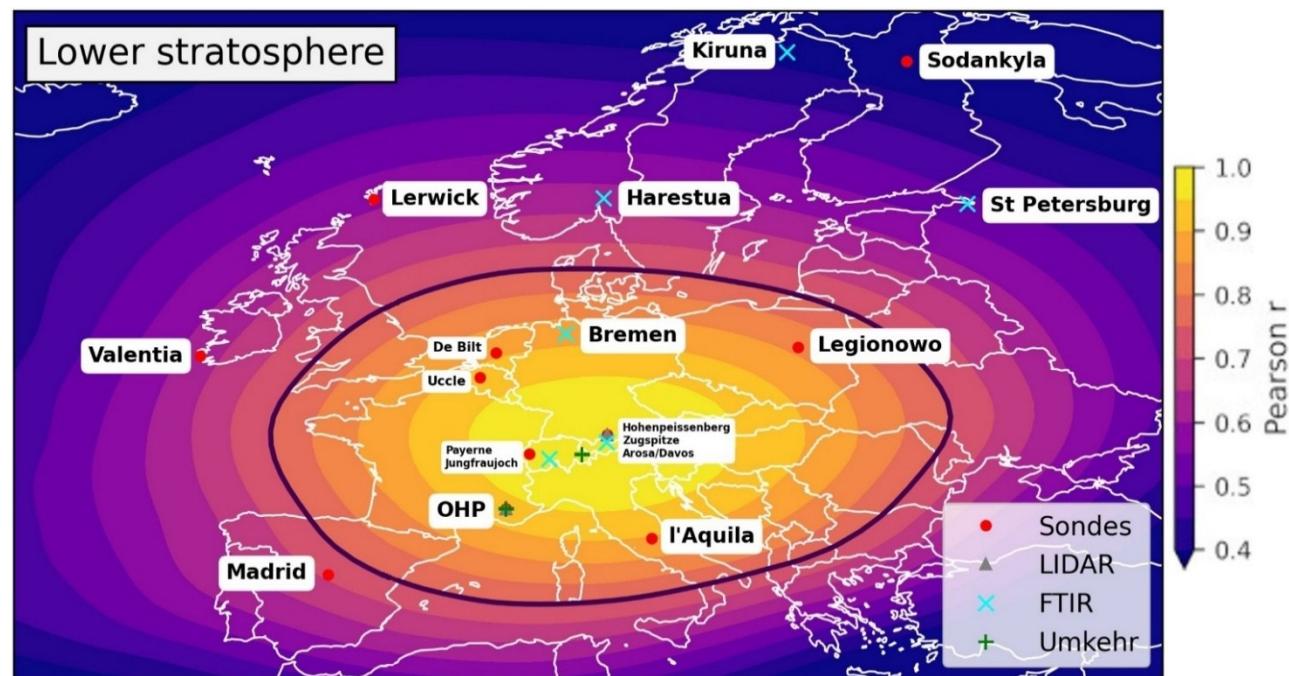



Figure 4: Correlation map for Zugspitze, FTIR (47.42∘ lat., 10.98∘ lon.) in the lower stratosphere. We see that all sites in the
310 Central Europe group lie within the r>0.75 limit (black line) of correlation with Zugspitze.

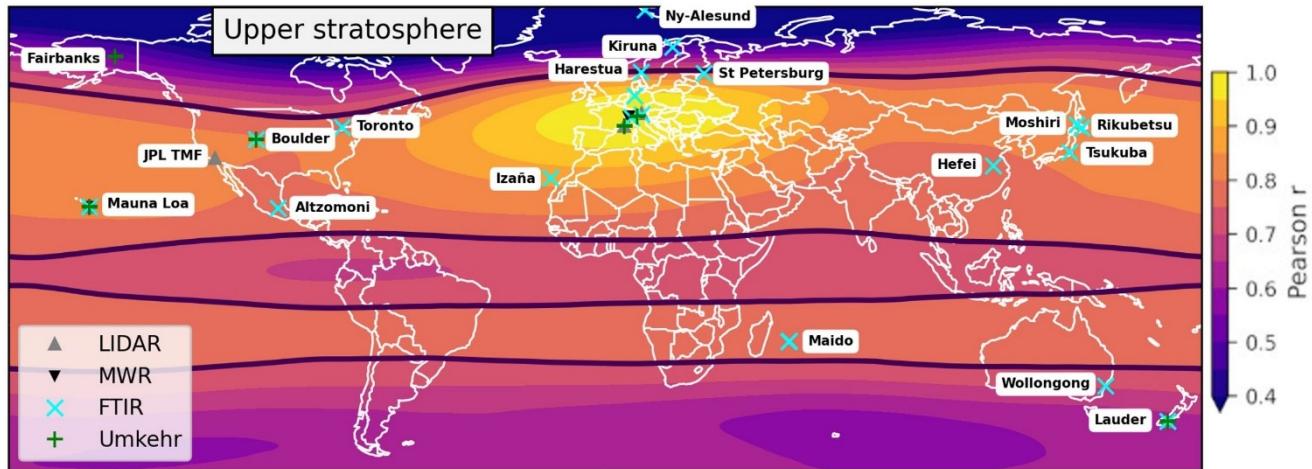



Figure 5: Correlation map for Zugspitze in the upper stratosphere. The r>0.75 region limit (black line) here spans the whole mid latitude region. In the upper stratosphere, the ozonesondes are not measuring so there are only 30 stations available there.

3.2 Composite time series

Merging heterogeneous records requires careful handling of systematic biases and outliers. In this section, we describe the pre-processing of the data records and compare two merging methodologies: the conventional weighted mean (WM) and the BAyeSian Integrated and Consolidated (BASIC) algorithm (Ball et al., 2017).

3.2.1 Offset removal

Systematic errors (including offsets) might exist between ozone datasets within the same region due to local atmospheric variability and differences between retrieval techniques. To account for this, following Ball's approach (Ball et al., 2017), we align all the time series to a common baseline. We calculate the mean value of a selected reference instrument (arbitrarily chosen) over an arbitrary reference period (2009–2014). All other series in the group are shifted (via subtraction of their respective means) to align with this reference. This process effectively removes relative offsets, resulting in a unified dataset of aligned monthly means that preserves the seasonality and long-term variability, suitable for trend analysis.

Mathematically, following the notation of Ball, for the full dataset $d_{t,c}$ ($n_c \cdot n_t$ matrix where n_c is the number of instruments and n_t is the number of measurements of one instrument), we define the aligned dataset $d'_{t,c}$, in which we removed the relative offsets, by adjusting the mean of each instrument c to the mean of the reference instrument $d'_{t,ref}$ over the reference period $[t_1, t_2]$:

$$d'_{t,c} = d_{t,c} - \frac{1}{n} \sum_{t'=t_1}^{t_2} d_{t',c} + \frac{1}{n} \sum_{t'=t_1}^{t_2} d_{t',ref} \quad (1).$$

Where n denotes the number of months between t_1 and t_2.

3.2.2 The Weighted Mean

335 One way of merging datasets is computing their WM. For a set of aligned monthly mean measurements x_i with uncertainties σ_i at a given time step t, over all the instruments c of the group, the WM is defined as:

$$WM(t) = \frac{\sum_c w_{t,c}\, d'_{t,c}}{\sum_c w_{t,c}}, \quad \text{with} \quad w_{t,c} = \frac{1}{\sigma_{t,c}^2} \quad (2).$$

The weights are the inverse of the squared measurement uncertainties (of L3 data in our case). The WM assumes that measurement errors are normally distributed and that all instruments provide unbiased estimates of ozone. However, the WM is highly sensitive to outliers. This is particularly problematic if an outlier instrument c at a month t has a small uncertainty:
340 the weight $w_{t,c}$ becomes disproportionately large, pulling the composite average toward the outlier value. The WM assumes the reported uncertainties result only of random normally distributed measurement errors. While the harmonization of individual datasets and our alignment of the time series supposedly removed systematic errors and biases that existed, the merging correlation threshold 0.75 accepts spatial variability, another source of inconsistency. Since the WM cannot distinguish between this spatial variability and the measurement uncertainties, it allows those outliers to bias the curve.

345 ### 3.2.1 The BASIC Methodology

BASIC (Ball et al., 2017) uses Bayesian inference (Rodgers et al, 2000) to determine the most probable underlying ozone time series by combining the data with prior knowledge of ozone variability and a robust handling of outliers.

The Bayesian method computes the posterior probability distribution of the true ozone time series y, given the observed dataset d and a month to month prior M as :

350
$$P(y|d, M) = \frac{P(y|M) \cdot P(d|y)}{P(d|M)} \quad (3)$$

$P(y|M)$ is the month-to-month prior distribution. It describes the expected change of ozone from one month to another. In Ball et al., the prior is computed from the actual datasets. Here, the prior is derived from the ML climatology (McPeters and Labow, 2012) consisting of Aura MLS and ozonesonde datasets, to ensure the independence of the prior towards the data records. As this climatology has a very large variability, the prior constraint of the optimal estimation method is small but
355 still ensures the convergence of the algorithm.

$P(d|y)$, the likelihood, represents the probabilistic model of the data. To account for outliers, BASIC uses a Gaussian-mixture model (Box & Tiao, 1968). Unlike a single Gaussian, it assumes that any data point has a probability β of being an outlier with an uncertainty inflated by a factor γ. This allows the algorithm to effectively down-weight data points that deviate significantly from the regional consensus. In this work, $\beta = 0.1$ and $\gamma = 100$, thus we expect 10% of outliers, that
360 can have a 100 times bigger variance.

Figure 6 illustrates this process for the Central Europe group in LS for the months of January, May and August 2012. The black curve images the probability density function of the observations, with the merged uncertainty (instruments plus PCA uncertainties, see 3.2.4), drawn as the sum of individual instrument Gaussian distributions. The green curve represents the prior distribution derived from the ML climatology- a very large distribution because of the high variability reported in the

365 climatology. BASIC samples the final posterior distribution P(y|M,d) using a Markov Chain Monte Carlo (MCMC) method. In May 2012 (Figure 6b), the instrument's curves are close, so the probability density is almost unimodal (Wilks et al., 2019). Therefore, the BASIC posterior (in blue) and the WM (in red) are very similar, BASIC being broader to account for the main large body of the distribution. In the image, the sum of instruments (black) curve is larger than the blue BASIC posterior because the black curve represents the sum of individual likelihoods, whereas the BASIC curve represents the product of the

370 likelihoods and the prior, which naturally sharpens the probability density around the consensus value. In January 2012 (Figure 6a), the datasets are disparate, and BASIC identifies the larger mode and ignores the outlier (peak visible at 145 DU), producing a robust estimate closer to the cluster of consistent instruments, whereas the WM is pulled toward the outlier. Finally, in August 2012 (Figure 6c), the probability density splits into two distinct peaks (near 93 DU and 103 DU). The Weighted Mean is centered at 97 DU —directly between the peaks— where the actual probability density is low. This

375 result is physically unlikely as it represents a value supported by fewer instruments than 102 DU (right peak) or 94 DU (left peak). In contrast, BASIC's Gaussian mixture model selects the more probable peak (centered at 94 DU), effectively resolving the ambiguity rather than averaging it out.

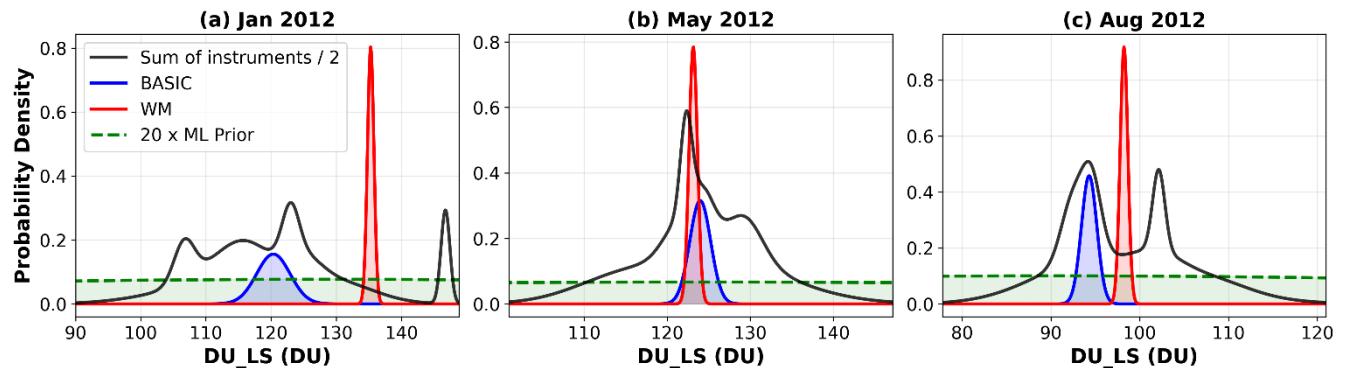



Figure 6: Illustration of the probability density sampling of BASIC, for January, August and May 2012, for the group Central
380 Europe in LS. The black distribution is the sum of all instruments' Gaussians, the green distribution is the ML prior distribution, the blue is BASIC and the red is the WM.

3.2.1 Principal Component Analysis (PCA) uncertainties

To further enhance the detection of outliers, we use a PCA to identify the most common ozone signal among all the time series. The beta coefficient of the probabilistic model (Box-Tiao) is set to 0.1, an estimation of the proportion of outliers that

385 becomes less robust as the number of instruments decreases. Our approach is inspired by Ball et al. (2017), except that we do

not use it to create the monthly mean uncertainties (as we propagate them from the L1 data) but to improve the outliers detection. A PCA is performed via Singular Value Decomposition on the aligned datasets d': the first mode represents the common underlying ozone signal, while the higher-order modes represent the noise and specific variability of each time series. Mathematically, the decomposition of d' is of the form $d' = U\,W\,V^T$, where U contains the temporal modes, W

390 contains the singular values (amplitude of each mode), and V the projection coefficients (how much each temporal mode in U is present in all the instruments). The statistical uncertainty $\sigma_{\text{PCA,t,i}}$ for instrument i at time t is obtained by summing the contributions of the noise modes (from k=2 to N), identically to equation 5 in Ball et al. (2017):

$$\sigma^2_{\text{PCA,t,i}} = \sum_{k=2}^{N} \left(U_{t,k}\, W_{k,k}\, V_{i,k} \right)^2 \quad (4)$$

Finally, to obtain the merged uncertainties, the instrumental L3 uncertainties (σ_{Profile}) generally have a much larger

395 magnitude than the statistical scatter (σ_{PCA}), so a direct sum is meaningless. To combine them, we use the PCA uncertainty as a scaling factor. For each month, we identify the "reference" instrument as the one with the smallest statistical uncertainty ($\sigma_{\text{PCA,ref}}$). The final merged uncertainty used in BASIC is given in Equation 5.

$$\sigma_{\text{Merge,i}} = \sigma_{\text{Profile,i}} \cdot \frac{\sigma_{\text{PCA,i}}}{\sigma_{\text{PCA,ref}}} \quad (5)$$

The resulting merged uncertainties contain in the end the L3 profile uncertainties, inflated with respect to their agreement to

400 the reference PCA uncertainty, which is the smallest PCA uncertainty for this month.

3.3 Multiple Linear Regression (MLR) trend estimation

Trends are estimated using the weighted multiple linear regression (MLR) model developed within the APARC/LOTUS initiative. This open-source regression tool (v0.8.3, see Supplement 5) has been widely used for evaluating stratospheric ozone profile trends since the WMO 2018 Assessment (WMO, 2018). Here we evaluate trends between January 2000 and

405 December 2024.

The multiple regression of the time serie y(t) is performed as follows:

$$y(t) = \beta_1(t) \cdot QBO_1(t) + \beta_2(t) \cdot QBO_2(t) + \beta_3 \cdot Solar(t) + \beta_4 \cdot NAO(t) + \beta_5(t) + \beta_6 \cdot t + \epsilon(t) \quad (6).$$

With $\beta_i(t) = \beta_{i,0} + \sum_{k=1}^{2} \beta_{i_1,k} \sin\left(\frac{2\pi k t}{12}\right) + \sum_{k=1}^{2} \beta_{i_2,k} \cos\left(\frac{2\pi k t}{12}\right)$ for $i = 1, 2, 5$.

410 The trend estimation is performed here directly on the monthly mean composite time series, and the annual cycle is modeled within the regression by expanding specific coefficients (the intercept $\beta_5(t)$ in Eq. 6) into Fourier series. Moreover, the merging methods provide uncertainties for their monthly mean values, which enable the use of a weighted MLR. The fit is constrained by the inverse of the squared BASIC uncertainties ($\frac{1}{\sigma^2_{BASIC}}$), so that months with smaller uncertainties – therefore higher confidence – contribute more to the trend result.

415

The other functions in Eq. 6 are the explanatory variables, which have been chosen to be the same for all the regional groups to make the inter-regional comparison easier:

- QBO: The Quasi-Biennial Oscillation is modeled using two orthogonal components (EOF-1 and EOF-2) derived from the Principal Component Analysis of Singapore stratospheric winds (10–100 hPa).
- 420 Solar Cycle: The 11-year solar cycle is represented by the 10.7 cm solar radio flux (F10.7).
- For Northern Hemisphere groups, we use the North Atlantic Oscillation (NAO) index. This is not the proxy used in previous LOTUS papers such as Godin-Beekman et al., 2022. A sensitivity study showed that the R^2_{adj} are similar using NAO or ENSO. Since most of the groups in this study are in the northern hemisphere, we selected the NAO proxy for all groups to ensure consistency across all groups.
- 425 For the Lauder (Southern Hemisphere) group, we replace NAO with the El-Niño-Southern Oscillation (ENSO) index (MEI v2). This proxy is used with the stratospheric Aerosol Optical Depth (sAOD) from GloSSAC v2.2 to account for volcanic perturbations (e.g., Raikoke, Hunga Tonga).

The linear trend is the coefficient β_6 in DU / month, which can then be converted into %/decade using the climatological
430 mean. The uncertainty of the trend estimates (σ_{Trend}) is derived from the standard error of the regression coefficient β_6. However, the autocorrelation in the fitting residuals ($\epsilon(t)$) typically leads to underestimated uncertainties. To correct this, we apply the Cochrane-Orcutt transformation (Cochrane-Orcutt, 1949) to remove first-order autocorrelation from the residuals. The quality of the regression fit is assessed using the Adjusted Coefficient of Determination (R^2_{adj}). Unlike the standard R^2, which increases automatically with the number of predictors, R^2_{adj} accounts for the degrees of freedom:

$$435 \qquad R^2_{adj} = 1 - (1 - R^2)(n - 1)/(n - p - 1)$$

where n is the number of observations and p is the number of predictors. A high R^2_{adj} (closer to 1) indicates that the chosen proxies and seasonal model successfully capture the observed ozone variability. On the other hand, a low R^2_{adj} suggests that the variability is driven by processes not included in the model (e.g., chaotic dynamical noise in the lower stratosphere), or that the instrumental noise exceeds the natural variability. Trends are considered statistically significant at the 95 %
440 confidence level if their absolute value exceeds their 2σ uncertainty (if |Trend| > 2 σ_{Trend}).

4 Results and Discussion

4.1 BASIC composites

The BASIC methodology was applied to all the regional groups of datasets that are described in Tables S2-S5 (see Supplement 2). Here, we display the group Central Europe as a representative case study to compare the performance of
445 BASIC with the conventional WM.

Figure 7 displays the time series in UpS. The aligned individual time series (top panel) show a seasonal cycle ranging generally between 40 and 60 DU, with individual outliers ranging from 30 to 70 DU. In this layer, there are no particular discrepancies between BASIC (in blue) and WM (in red) composites (bottom panel). The differences between the two methods are minimal, indicating that when the underlying data are consistent, either composite method converges to the standard result.

In Figure 8, corresponding to MS, the annual variability ranges from 80 to 110 DU. Divergences between the merging methods are visible, e.g. in 2005–2006, where the WM is consistently higher than BASIC during almost 2 years. This is driven largely by the O3S Legionowo (and other ozonesondes), which are significantly higher than the other instruments during this period. BASIC remains centered on the majority of the instruments, but with a larger uncertainty to include the larger range of variability. In 2017-2018, the opposite occurs: O3S Legionowo (and other ozonesondes) are lower than the rest, dragging the WM downward. Finally, during the beginning and the end of the time series, the instrumental variability is large, and BASIC uncertainties reflect the lack of consensus by being larger. In contrast, the WM uncertainties remain unrealistically small, failing to capture the structural variability at this month.

In Figure 9, in LS, we observe the largest dispersion, ranging from 90 to 175 DU, with outliers extending from 60 to 210 DU. While the BASIC and WM time series often overlap, their uncertainty characterization differs fundamentally. Figure 10 compares individual instrument monthly means with the BASIC and WM composites and their $\pm 2\sigma$ uncertainty intervals. Mathematically, the BASIC algorithm assumes a Box–Tiao error distribution (with outlier proportion $\beta=0.1$ and variance inflation $\gamma=100$). In such a distribution, one expects 92% of the data to be included within the $\pm 2\sigma$ uncertainty intervals, and 99.7% of the data for a pure Gaussian. In Figure 10a (February 2012), the instrument values are widely scattered (105–165 DU). The BASIC algorithm accounts for this dispersion with a large posterior uncertainty: the BASIC $\pm 2\sigma$ interval includes 5 out of 12 instruments. The WM $\pm 2\sigma$ interval however is extremely narrow, overlapping with only 1 of 12 instruments. In Figure 10b (September 2012), where agreement is better (80–110 DU), the BASIC uncertainty interval is smaller and overlaps with 8 of 12 instruments. The WM $\pm 2\sigma$ interval overlaps with 2 of 12 instruments. These results are to be interpreted cautiously because the time series are correlated and biases may exist. These coverage counts are only illustrative; however, they clearly highlight the robustness of the Bayesian error inflation to deal with outliers compared to the conventional WM.

Finally, Figure S1 presents the Tropospheric partial column. Despite periods of high instrumental variability (e.g., March 2004, where the BASIC uncertainty exceeds 10 DU to account for the inter-instrument spread), the overall time series of BASIC and the WM overlap significantly. This agreement confirms that BASIC is robust enough to retrieve the common geophysical signal even in the most heterogeneous layer of the atmosphere. The distinct advantage of BASIC lies in its uncertainty characterization: it accounts for periods of low consensus with large uncertainties, whereas the WM ignores it.

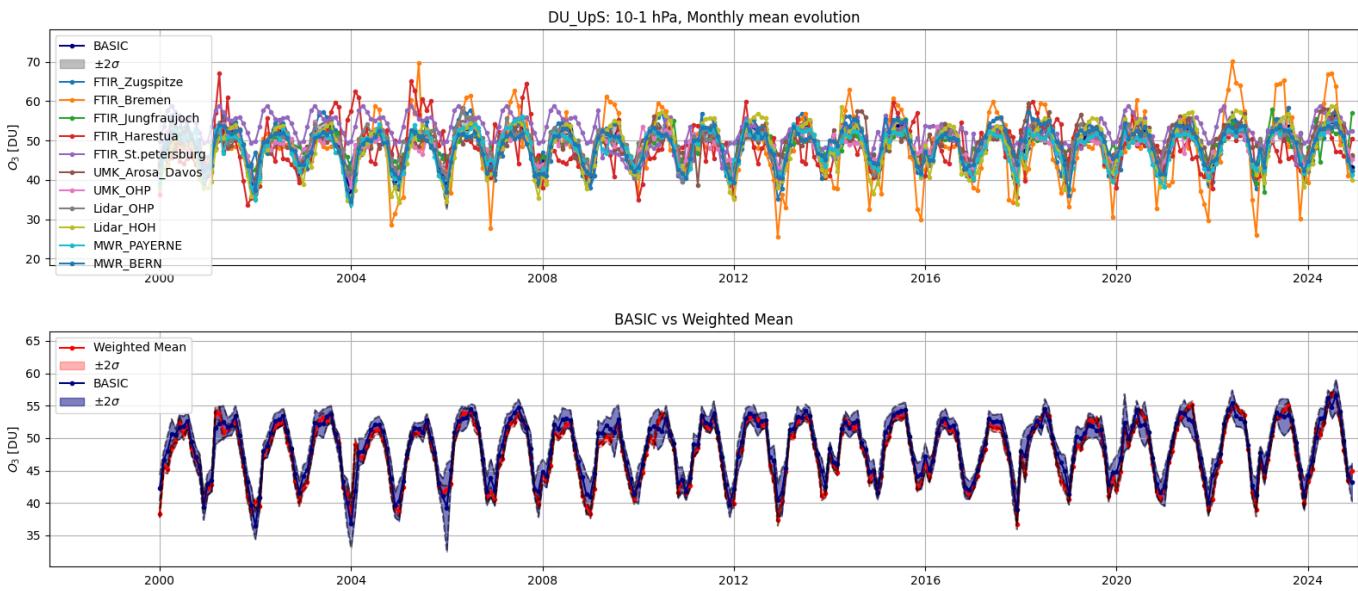



Figure 7: Time series for regional group central Europe, in the Upper Stratosphere (UpS). Top: time series of the group
480   members, plotted with BASIC in blue and its uncertainty in grey. Bottom: comparison between BASIC in blue, with 2
uncertainties in blue, and the Weighted mean in red, with 2 uncertainties in red.

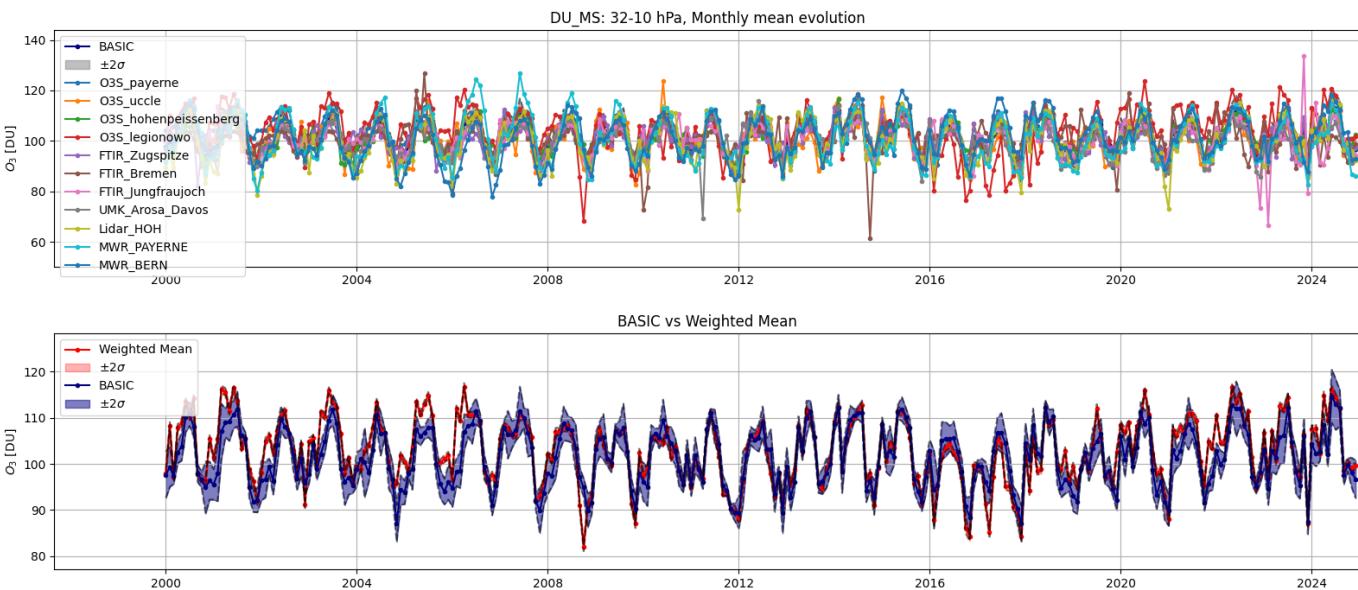



Figure 8: Same as Fig. 7 for the Middle Stratosphere (MS).

485

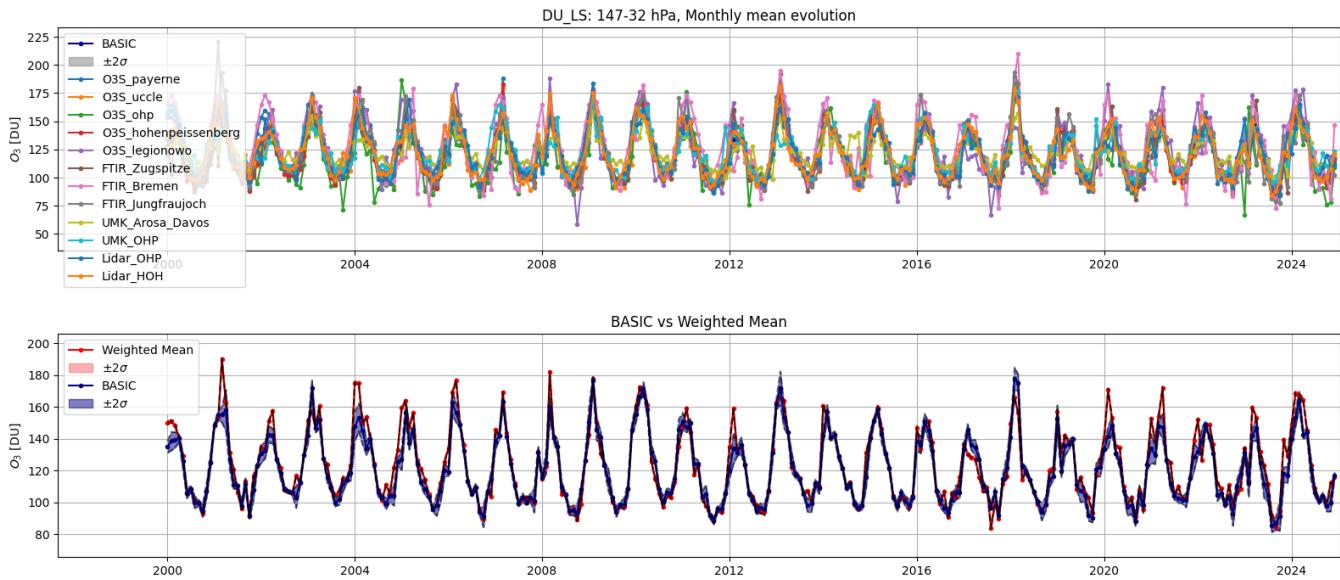



Figure 9: Same as Fig. 7 for the Lower Stratosphere (LS).

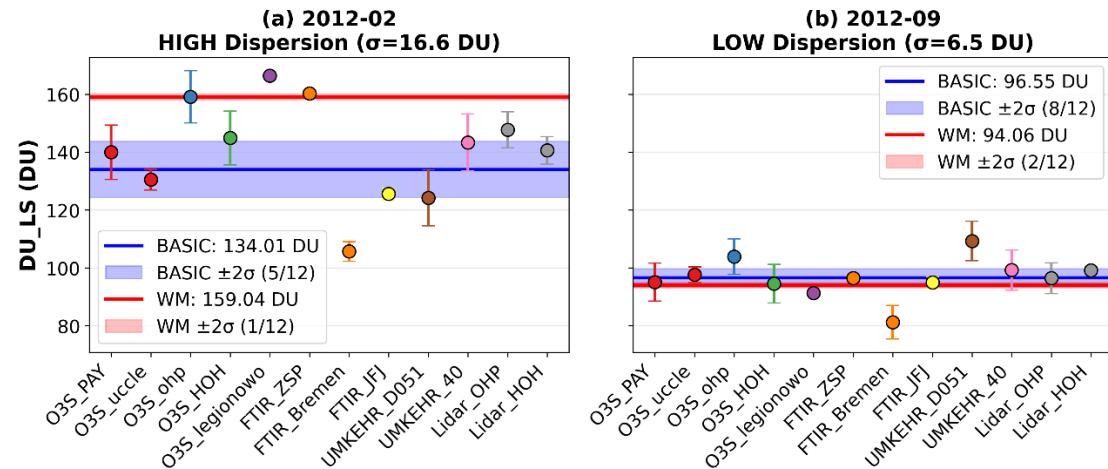



Figure 10: Monthly mean ozone values in DU for all the instruments in the group Central Europe in LS, for two months selected for their high (left panel, Feb. 2012) and low (right panel, Sept. 2012) variability. The BASIC (in blue) and WM (in red) composites are superposed with their 2 uncertainties.

4.2 Vertical regional trends

The trends for the 2000-2024 period are estimated using the LOTUS version 0.8.3 regression model described in Section 3.3. We illustrate the results here for three groups of selected NDACC stations: the mid-latitude Central Europe, Mauna Loa (MLO) in the tropics, and Lauder in the Southern-Hemisphere, as investigated in Godin-Beekmann et al. (2022) for the period 2000-2020 and Sofieva et al. (2025) for the period 2000-2024, but results are provided for all regions and all layers in

Sect. 4.3. In the figures of this section (e.g. Figure 11), the legend lists the datasets for which the trends have not been estimated on the same time range as the composite trends period, either beginning later than 2000, or finishing before 2024, and/or who have significant data gaps. The complete list of time range and gaps for each instrument is available in Table S1.

500 A common misconception is to consider that the weighted mean trend is a weighted mean of the trends of the individual instruments. It would hold only if the commutator between the operator of merging (BASIC or the weighted mean) and the operator of trend estimation (MLR) would be 0. Because of the reduction of dimensionality in the merging techniques (from n timeseries to 1) and in the MLR (from more than 220 monthly mean to 2 values), this commutator is likely different from 0, which means the trends from the composites cannot be considered as the mean of the individual trends. We chose to

505 include individual trends to illustrate the reduction of the uncertainty allowed by the merging techniques.

In table 2, we display the estimated trends for the two sets of partial columns oPC and aPC, for the two merging techniques and for three selected regions (Europe, Hawaii and Lauder). BASIC trend uncertainties are systematically smaller than those of the WM (except in the group Hawaii in TROPO). Over these groups, BASIC uncertainties are on average 15.3% smaller than the WM's ($= \frac{\sigma_{BASIC} - \sigma_{WM}}{\sigma_{WM}}$). For Central Europe, we compute an average reduction of 15 % in oPC, 21 % in aPC, for

510 Hawaii, 17 % in oPC and aPC, and for Lauder 15 % in oPC and 7 % in aPC.

oPC	TROPO	$2\sigma_{\text{TROPO}}$	LS	$2\sigma_{LS}$	MS	$2\sigma_{MS}$	UpS	$2\sigma_{UpS}$
Central Europe BASIC	-1.14	1.67	-1.07	1.09	0.63	0.69	**1.44**	**0.55**
Central Europe WM	-1.1	2.07	**-1.81**	**1.33**	-0.1	0.95	**1.77**	**0.53**
Hawaii/MLO BASIC	**2.30**	**2.15**	1.09	2.17	**-0.89**	**0.60**	1.02	0.49
Hawaii/MLO WM	1.69	2.07	2.05	2.44	**-1.25**	**0.93**	1.89	0.66
Lauder BASIC	0.74	1.61	**-2.12**	**1.53**	**-1.61**	**0.95**	1.61	0.75
Lauder WM	0.60	1.77	-1.67	1.68	**-1.45**	**1.08**	1.38	1.07
aPC	aTROPO	$2\sigma_{\text{aTROPO}}$	UTLS	$2\sigma_{UTLS}$	aMS	$2\sigma_{aMS}$	aUpS	$2\sigma_{aUpS}$
Central Europe BASIC	-0.09	0.94	**-2.59**	**1.85**	**-1.22**	**0.87**	0.31	0.48
Central Europe WM	-0.13	1.79	**-3.52**	**2.22**	-0.53	1.05	**0.80**	**0.51**
Hawaii/MLO BASIC	**2.48**	**1.92**	2.68	2.85	-0.39	1.45	0.37	0.42
Hawaii/MLO WM	**2.02**	**2.01**	2.48	3.56	1.26	1.77	0.49	0.58
Lauder BASIC	1.14	1.59	-1.03	2.24	**-1.30**	**1.16**	0.51	0.89
Lauder WM	0.59	1.68	-0.09	2.36	**-1.93**	**1.15**	-1.00	1.11

Table 2: BASIC and WM trends and trend uncertainties in % per decade for three selected NDACC stations, for oPC and aPC. In bold are significant trends and their uncertainty.

4.2.1 Central Europe

515 Figures 11 and 12 display the trends over Central Europe for the two sets of partial columns, oPC and aPC respectively. While individual dataset trends show substantial scatter (considering only the instruments' trends with complete period, see Table S1 for data availability), the BASIC and WM trend composites generally align. The adjusted R2 values (Figure 12) are systematically higher for the composite trends than for individual record trends, therefore justifying the merging/composite approach.

520 In the Upper Stratosphere (UpS, 10–1 hPa), trends are positive and significant (+1.44 ± 0.55 %/decade). The Middle Stratosphere (MS) shows a positive non-significant trend. On the other hand, trends in the Lower Stratosphere (LS) and the Troposphere (TROPO) are negative and non-significant. The trend uncertainties for the BASIC composite are systematically smaller than those of the WM and of individual datasets.

For the alternative partial columns (aPC, Fig. 12), in the aUpS, BASIC (+0.31 ± 0.48 %/dec, non-significant) and the WM 525 (+0.80 ± 0.51 %/dec, significant) have a different trend significance, though BASIC gives smaller uncertainties. In UTLS, both methods reveal a strongly negative and significant trend (magnitude > -2.5 %/dec). The aPC definition allows the variability of the UTLS signal to be isolated in a single layer, while this is smoothed out in the standard LS partial column. Finally, the aTROPO trends are near zero and non-significant.

Our results align well with Sofieva et al. (2025), who report trends over the period 2000-2024 for Alpine Stations (Central 530 Europe) in their Figure 6c. Note that in Sofieva et al. (2025), the merging is done after the trend estimation and that we compare highly resolved vertical trends to partial column trends. We report a transition from significant positive trends in the UpS to significant negative trends in the LS/UTLS, similar to Sofieva et al. (2025), Godin-Beekmann et al. (2022) and Petropavlovskikh et al. (2025). In UTLS/LS, we report a significant negative trend in the UTLS (-2.59 ± 1.85 %/decade) confirmed by the negative trend in Sofieva et al. (2025) and Godin-Beekmann et al. (2022). The inversion height can be 535 defined by the point where the negative trends become positive (or vice versa). Petropavlovskikh et al. (2025) suggest an inversion height in the 20–10 hPa region. Our data supports this, as the aMS (64–20 hPa) is significantly negative while the MS (32–10 hPa) is positive. Finally, we report a non-significant trend in the troposphere while Van Malderen et al. (2025b), report weakly positive trends. Note that the trend estimation method, the considered regions and the merging method are different in both studies, and we have considered the estimation of tropospheric trends to be beyond the scope of this 540 study. However, we can say that the slightly more negative trend estimate in the standard TROPO than in the aTROPO suggests that the TROPO layer may be capturing some of the negative signal from the lower stratosphere.

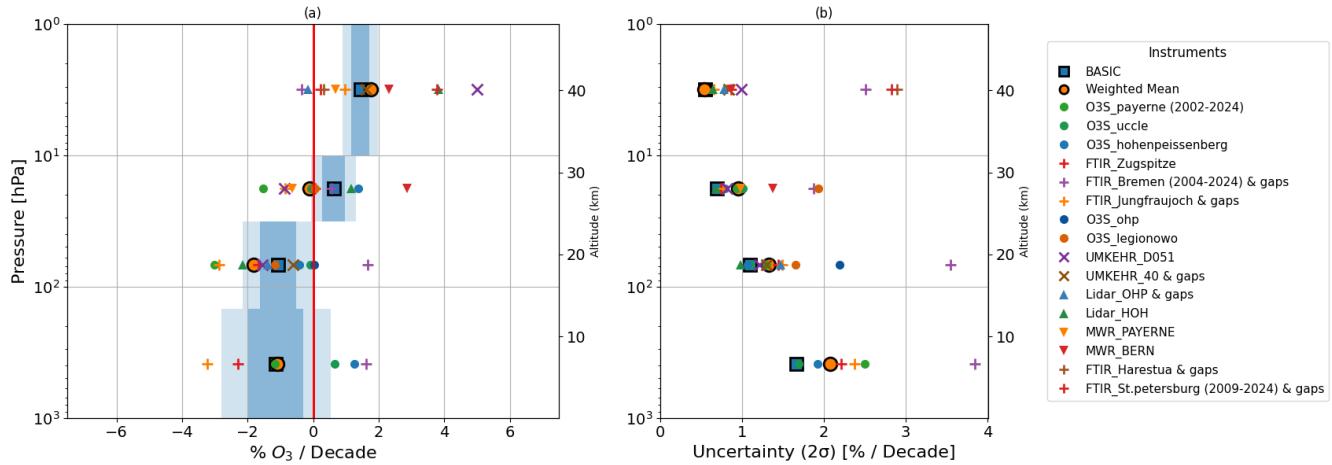



Figure 11: Trend estimates for the oPC of the regional group "Central Europe" in (a) 2000-2024 trends in % per decade. BASIC trend values are a blue square with black contour, with 1 sigma uncertainty in light blue shading and 2 sigma in darker blue shading. The WM trend is an orange circle with black contour. Trends for each individual dataset involved in the merged composite are plotted for comparison. (b) 2-sigma uncertainties of the composites and of each dataset are shown in % per decade. The legend highlights datasets with time ranges deviating from the 2000-2024 time period and/or with substantial gap.

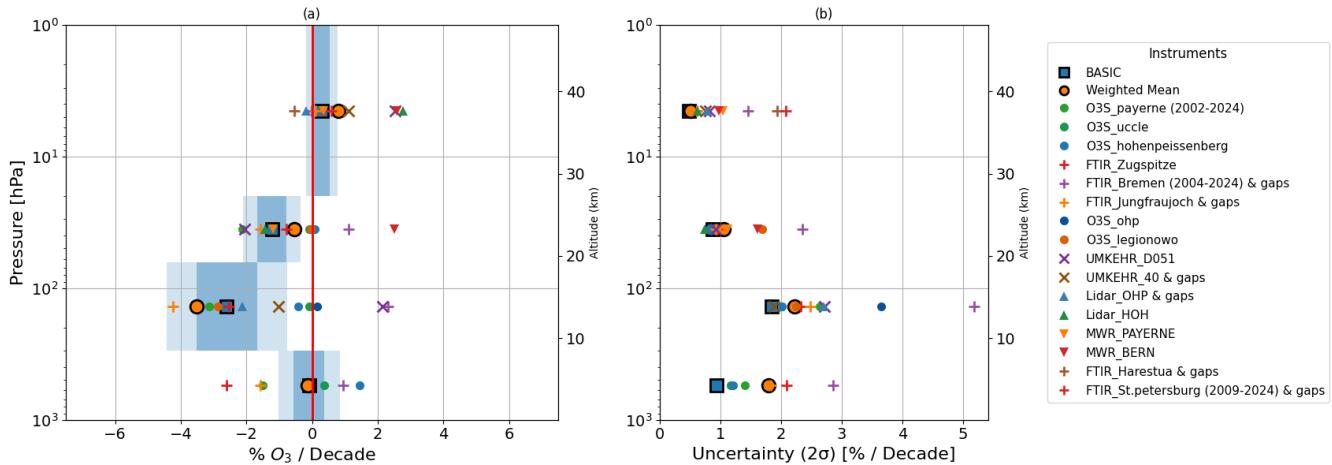



Figure 12: Same as 11 for the aPC of the group "Central Europe".

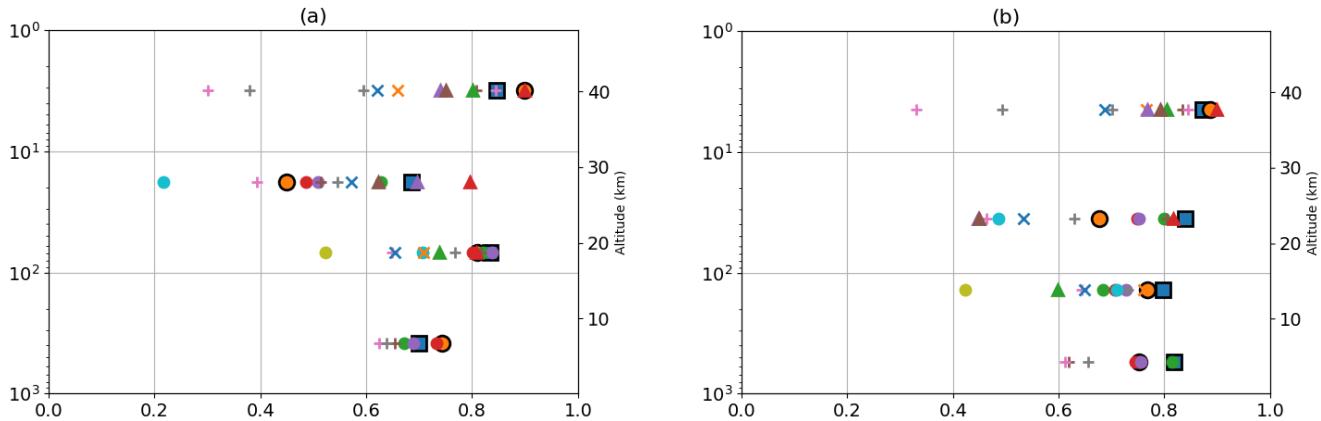



Figure 13: Adjusted R2 for BASIC, the weighted mean and the datasets of the Central Europe group for (a) oPC and (b) aPC.

4.2.2 Hawaii / Mauna Loa (MLO)

Figures 14 and 15 display the trends for the Hawaii region (see Table 2 for values). The UpS partial column is the broadest
555 "Mid-Latitudes" group (24+ instruments), while lower partial columns rely on the MLO specific datasets (disparate geographical spread, see table S2-S5).

In the Upper Stratosphere (UpS in oPC), both the BASIC and WM methods reveal statistically significant positive trends (BASIC: +1.02 ± 0.49 %/decade). On the other hand, in MS (oPC), both composites show statistically significant negative trends (BASIC: -0.89 ± 0.6 %/decade). In LS (oPC), the observed near zero trends lack statistical significance (BASIC:
560 +1.09 ± 2.17 %/decade) with a large uncertainty explained by low adjusted R2<0.5, see Figure 16. Although none of the single instruments gives a significant negative trend, it is worth noting that the FTIR in LS has a positive trend, but the time series stops in 2022. Finally, BASIC demonstrate statistically significant positive trends in the troposphere (BASIC: +2.3 ± 2.15 %/decade). In a nutshell, BASIC composite exhibits systematically smaller uncertainties across all partial columns than the WM, except for the standard TROPO where they are comparable but smaller for the WM (BASIC 2.15 %/dec vs. WM
565 2.07 %/dec).

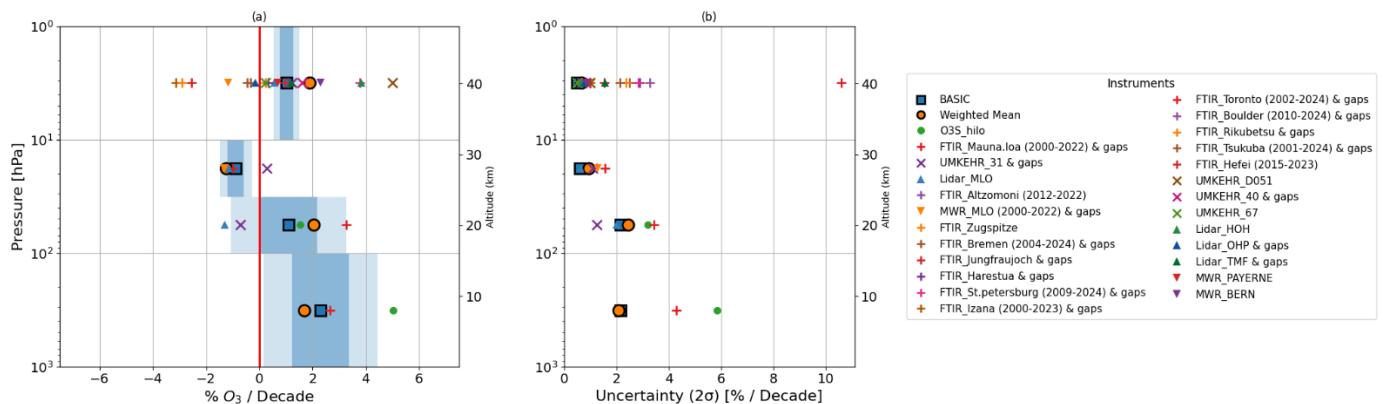
For the alternative partial columns (aPC), discrepancies arise in aMS, where BASIC estimates a negative trend (-0.39 ± 1.45 %/decade) while the WM suggests a positive one, though neither is statistically significant and both exhibit low adjusted R2 (< 0.6). The individual instruments are diverging from the very positive FTIR to the very negative MWR, both with incomplete period (2000-2022) and gaps. In UTLS, both methods indicate positive trends (BASIC: +2.68 ± 2.85 %/decade),
570 but these are associated with large uncertainties and low adjusted R2 (< 0.6), though better than in LS. In contrast, the UpS and aUpS layers exhibit a high trend reliability with adjusted R2>0.8.

Our results diverge from the literature in the layers below UpS. Sofieva et al. (2025) report slightly positive UpS trends transitioning to negative trends in the tropical UTLS. In contrast, our analysis gives positive non-significant UTLS trends. This discrepancy may be partially attributed to the inclusion of the Mauna Loa FTIR record in our analysis, which is absent

575 from the Sofieva et al. study. The FTIR record is temporally sparse (with significant gaps in 2001–2005 and 2011) and leads to unreliable individual trend estimates, which was the main reason on why it was not used in Sofieva et al. (2025). However, the BASIC methodology allows us to retain its physical information while properly weighting its uncertainties, potentially modifying the regional signal in the LS/UTLS. Furthermore, our observation of significantly negative trends in the MS contrasts with the generally zero or positive trends expected in the subtropics. This negative signal is likely

580 influenced by the ozonesonde records, as noted by Sofieva et al. (2025). Note that the HILO ozonesonde time series shows a total column ozone drop-off since approximately 2015. This decline is described in literature as an instrumental artifact rather than a physical atmospheric change (Stauffer et al., 2020; Stauffer et al., 2022).

These discrepancies may also come from the choice of explanatory variables. Hawaii is physically located in the ENSO-dominated Pacific, but we used the NAO proxy to maintain methodological homogeneity across the Northern Hemisphere

585 groups. Sensitivity tests replacing NAO with ENSO (see Appendix A) reveal that for the aPC set, stratospheric trends become nearly zero, while the aTROPO trend remains strongly positive. But the limited R2 differences between the 2 proxies choices indicates that a deeper investigation is needed, beyond the scope of this study.

590 Figure 14: Trend estimates for the oPC of the regional group Hawaii. (a) 2000-2024 trends in % per decade. BASIC trend values are a blue square with black contour, with 1 sigma uncertainty in light blue shading and 2 sigma in darker blue shading. The WM trend is an orange circle with black contour. Trends for each individual dataset involved in the merged composite are plotted for comparison. (b) 2-sigma uncertainties of the composites and of each dataset are shown in % per decade. The legend highlights datasets with time ranges deviating from the 2000-2024 time period and/or with substantial

595 gap.

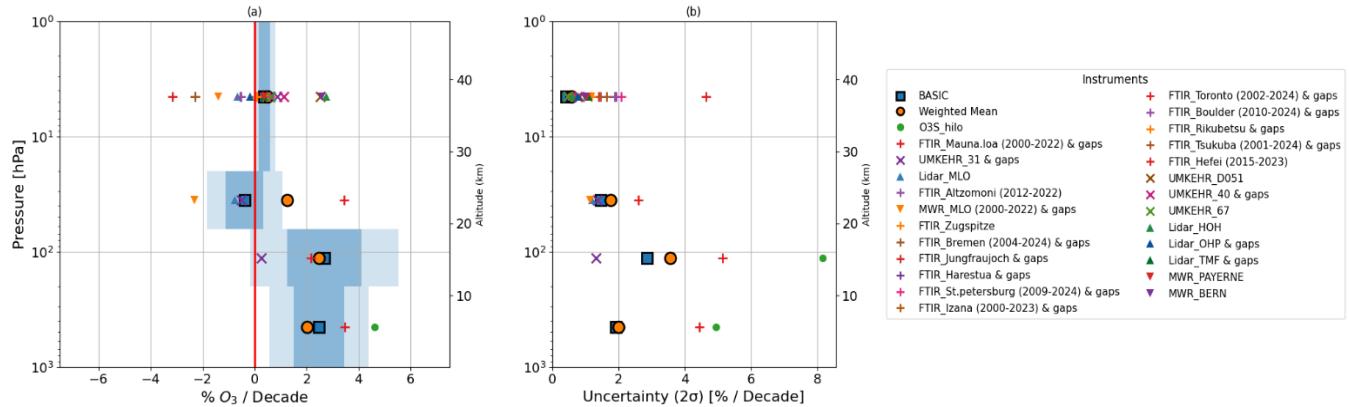



Figure 15: Same as 14 for the aPC of the group Hawaii.

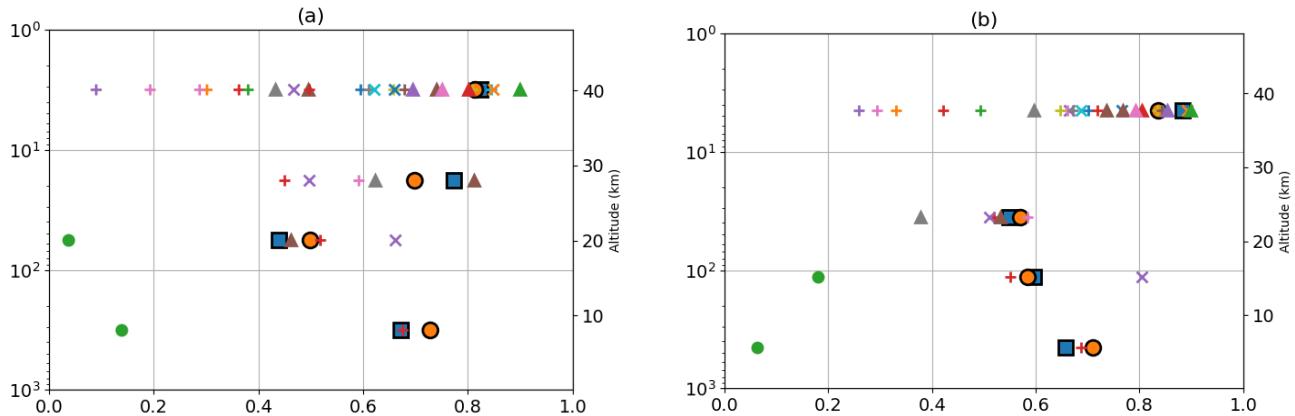



Figure 16: Adjusted R2 for BASIC, the weighted mean and the datasets of the Hawaii group for (a) oPC and (b) aPC.

4.2.3 Lauder

Figures 17 and 18 display the trends for the Lauder group (values in Table 2), estimated using ENSO proxy instead of NAO since this group is in the Southern Hemisphere.

In the UpS, both composites show statistically significant positive trends, with BASIC estimating a recovery of $+1.61 \pm 0.75$ %/decade. Similarly, in the MS and LS, the trends are negative and statistically significant across both methods (BASIC LS: -2.12 ± 1.53 %/decade; MS: -1.61 ± 0.95 %/decade). The robustness of the BASIC composite in these layers is further supported by the adjusted R2 values shown in Figure 19, which are consistently higher for the merged dataset compared to individual records.

However, disagreement emerges in the UTLS region: the BASIC composite indicates a negative, though non-significant, trend (-1.03 ± 2.24 %/decade, non-significant), whereas the WM suggests a near-zero trend (-0.09 ± 2.36 %/decade, non-significant). Our results in the lower stratosphere are highly consistent with Björklund et al. (2024) and Zeng et al. (2024),

who both report significant negative trends for the 2000–2020 period in LS (maximal near 20km in Björklund et al., 2024). The transition from positive trends to negative trends in the upper to middle stratosphere is further confirmed by our aUpS

615   (20–1 hPa) trend (+0.51 ± 0.89 %/decade, non-significant), which is statistically non-significant and near zero, aligning with Zeng et al. (2024) finding of a shift from positive trends to negative trends above 30 km.

A trend distinction emerges in the LS compared to Sofieva et al. (2025). While they report negative but non-significant trends for this region, our BASIC composite produces significant negative trends in the LS with high reliability (adjusted $R^2 > 0.8$ in Figure 19). Since the WM is also negative and significant here, it suggests that it is the partial column definition

620   that primarily enables the detection of a significant signal, while other vertical ozone profiles do not give significant trends. Finally, in the troposphere, both the standard (TROPO) and alternative (aTROPO) PC reveal positive but non-significant trends (+0.74 ± 1.61 %/decade and +1.14 ± 1.59 %/decade, respectively). These results align with those reported by Van Malderen et al. (2025a,b), Zeng et al. (2024) and Björklund et al. (2024). The slightly more positive trend in the aTROPO compared to the standard TROPO is consistent with the trend profile shape described by Zeng et al. (2024), where positive

625   trends peak near 5 km, and suggests the influence of the UTLS in the layer TROPO, pulling towards more negative trends.

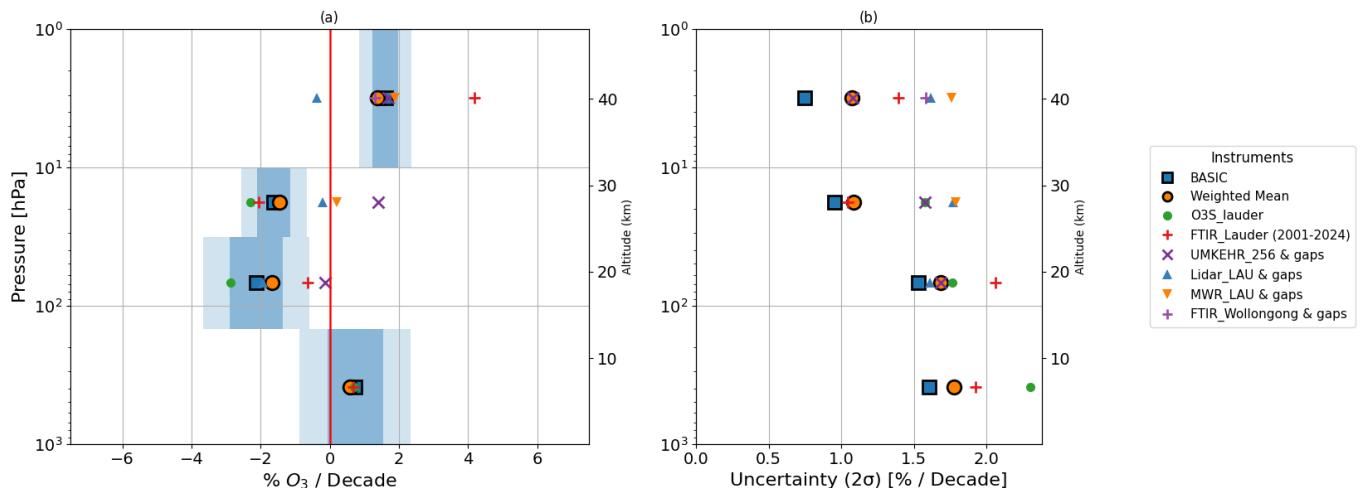
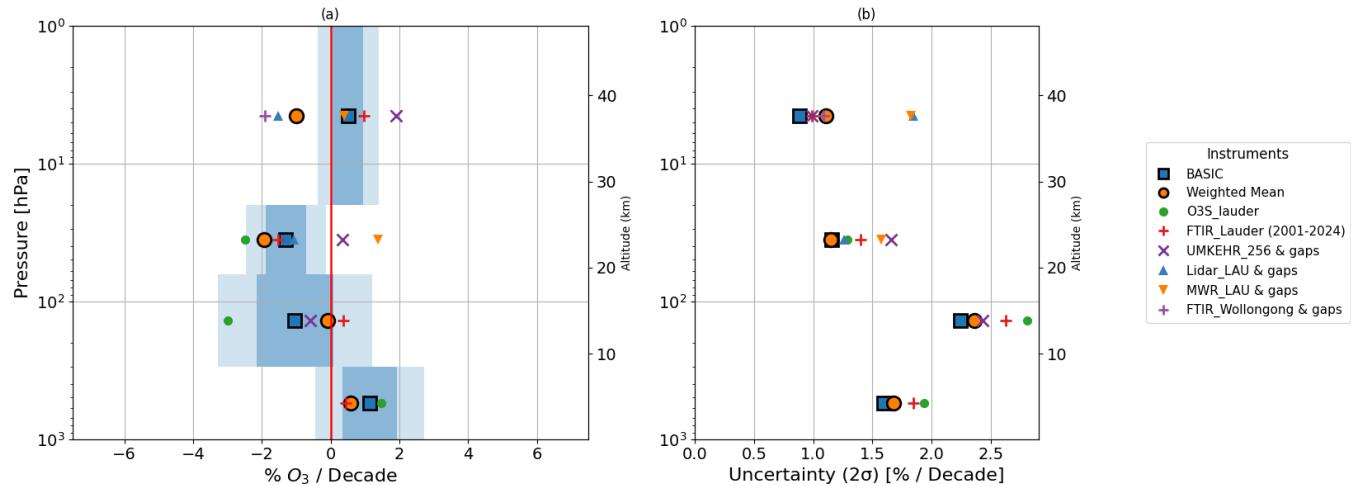



Figure 17: Trend estimates for the oPC of the regional group Lauder. (a) 2000-2024 trends in % per decade. BASIC trend values are a blue square with black contour, with 1 sigma uncertainty in light blue shading and 2 sigma in darker blue

630   shading. The WM trend is an orange circle with black contour. Trends for each individual dataset involved in the merged composite are plotted for comparison. (b) 2-sigma uncertainties of the composites and of each dataset are shown in % per decade. The legend highlights datasets with time ranges deviating from the 2000-2024 time period and/or with substantial gap.

635   Figure 18: Same as 17 for the aPC of the group Lauder.

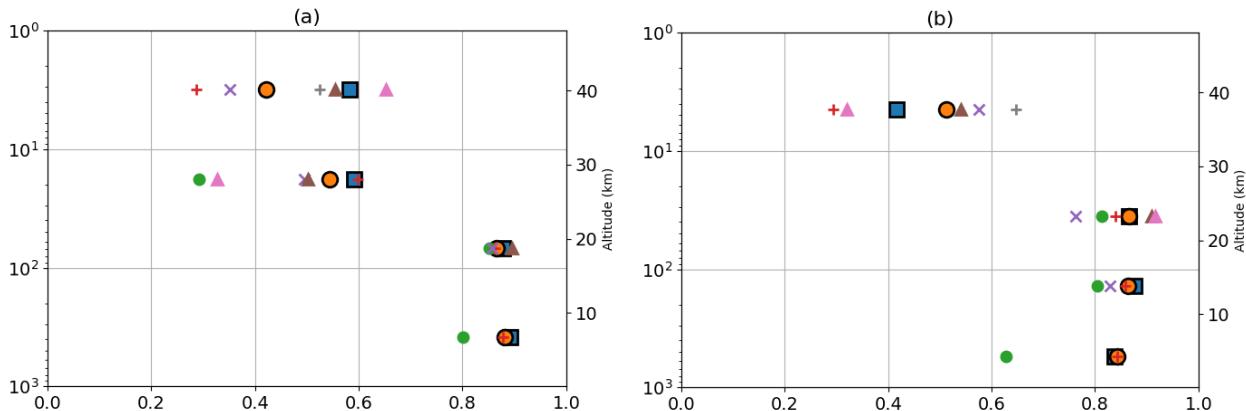



Figure 19: Adjusted R2 for BASIC, the weighted mean and the datasets of the Lauder group for (a) oPC and (b) aPC. Legends are similar to those of figures 17 and 18.

4.3 Global maps of trends

To visualize the spatial distribution of ozone recovery, we display global maps of trend estimates for all the partial columns defined above. Figures 20–22 and S2 show trends for the original oPCs, and Figs. 23, S3-S5 show trends for the alternative aPCs.

The maps only display trends derived from BASIC-merged regional groups and not individual instruments' trends. This choice follows the methodology in order to maximize statistical significance and minimize trends uncertainty. However, it results in gaps in small geographical coverage—like in South America, Southern Africa, and Siberia—where the NDACC network density is insufficient to form groups.

In UpS (Figure 20), trends are generally positive and significant across the globe. We estimate robust recovery of +1.0 ± 0.5 %/decade in the Mid-Latitudes, +1.4 ± 0.5 %/decade in Central Europe, and +1.6 ± 0.8 %/decade in Lauder. In the alternative aUpS (20–1 hPa equivalent to 30-40 km, Figure S3), the amplitude of these trends decreases, and they are non-significant for most regions (e.g., Mid-Latitudes +0.4 ± 0.4 %/decade; Central Europe +0.3 ± 0.5 %/decade). Scandinavia has a positive trend (+0.6 ± 1.5 %/decade). North Canada shows opposite trends in aUpS (positive, 1.3 ± 2.4 %/decade) than in UpS (negative, -1.0 ± 2.7 %/decade), underlying the limitations of BASIC for sparse datasets. Because of the polar night, the FTIR (only instrument in the group north Canada in UpS and aUpS) cannot measure during winter. The more important daily cycle (10 % near 50 km, according to Schranz et al., 2028) and the sensibility to NOX (importance if the choice of proxies) suggests that our analysis here is not reliable.

In MS (25-30 km, Figure 21), trends are negative and significant near the equator. The European continent, however, displays a mixture of positive and negative non-significant trends: Central Europe group (+0.6 ± 0.7 %/decade), North Europe (-1.1 ± 1.2 %/decade), South Europe (+0.3 ± 0.6 %/decade), and Svalbard/Greenland (-0.9 ± 1.8 %/decade). This points towards an underlying trends uncertainty in Europe, although it contains the densest networks of ground-based instruments. Strong significant negative trends are observed in the Equator and South Atlantic groups, as well as in Hawaii and the West US (-2.3 ± 0.7 %/decade). South Canada and North Canada show positive non-significant trends. These results generally agree more with the 25 km trend maps in Figure 7 from Sofieva et al. (2025) than with the 30 km maps, consistent with negative trends in the lower part of the layer and a transition from positive to negative near 30km altitude. In aMS (20-25 km, Figure S4), trends are consistently negative and non-significant, with few exceptions like Central Europe (-1.2 ± 0.9 %/decade) and East Coast (-7.7 ± 2.7 %/decade), again consistent with the 25km maps from Sofieva et al. (2025), which also reports from negative non-significant trends in the mid latitudes to weakly non-significant trends at 45° north and south. The East Coast US group shows a large significant decline, coherent with the strong depletion seen in the SCIAMACHY-OMPS 20 km maps (Sofieva et al., 2025).

In LS (15-25 km, Figure 22), trends are mostly negative and non-significant, with even a significant negative trend at Lauder (already discussed in section 4.2). Positive (non-significant) signals are seen in the West US (+0.7 ± 1.3 %/decade), Hawaii (+1.1 ± 2.2 %/decade) and South Canada (+1.0 ± 4.3 %/decade), contrast with the negative significant trends in East Coast (-3.3 ± 3.1 %/decade). This widespread negative trend is consistent with satellite measurements at 20 km (middle of LS) reported by Sofieva et al. (2025). In UTLS (10-20 km, Figure 23), we find significant negative trends in the Equator (-2.9 ± 2.2 %/decade), North Sea (-2.5 ± 2.3 %/decade) and Europe (-3.5 ± 2.3 %/decade), consistent with the 20 km satellite maps. However, an exception appears over North America: the groups West US (+1.8 ± 1.6 %/decade) and East Coast US (+3.9 ± 2.7 %/decade) show significant positive trends. This contradicts the negative trends from satellite maps at 20 km. However, the UTLS partial column has a large vertical range (approx. 10–20 km), which includes part of the upper troposphere. From TOAR-II analyses (Van Malderen et al., 2025b) confirmed by Thompson et al. (2025), we expect positive tropospheric trends in these regions, so our UTLS result likely reflects an aggregation of two opposite evolutions: a positive trend in the upper troposphere masking the negative trend in the lower stratosphere. The OCTAV-UTLS findings (Millàn et al., 2025)

underline the idea that the trends in pressure coordinates often combine signals from the troposphere and from the stratosphere, resulting in masking of the underlying trends.

Finally, trends in the standard troposphere (Figure S2) are generally non-significant. An exception is the North Canada
685 group, which shows a strong negative trend (-7.6 ± 3.8 %/decade), a decline confirmed by the findings of Van Malderen et al. (2025b) for the Canadian Arctic region. In aTROPO (Figure S5), for the same North Canada group, we also find a slightly negative trend, although non-significant (-1.1 ± 2.7 %/decade).

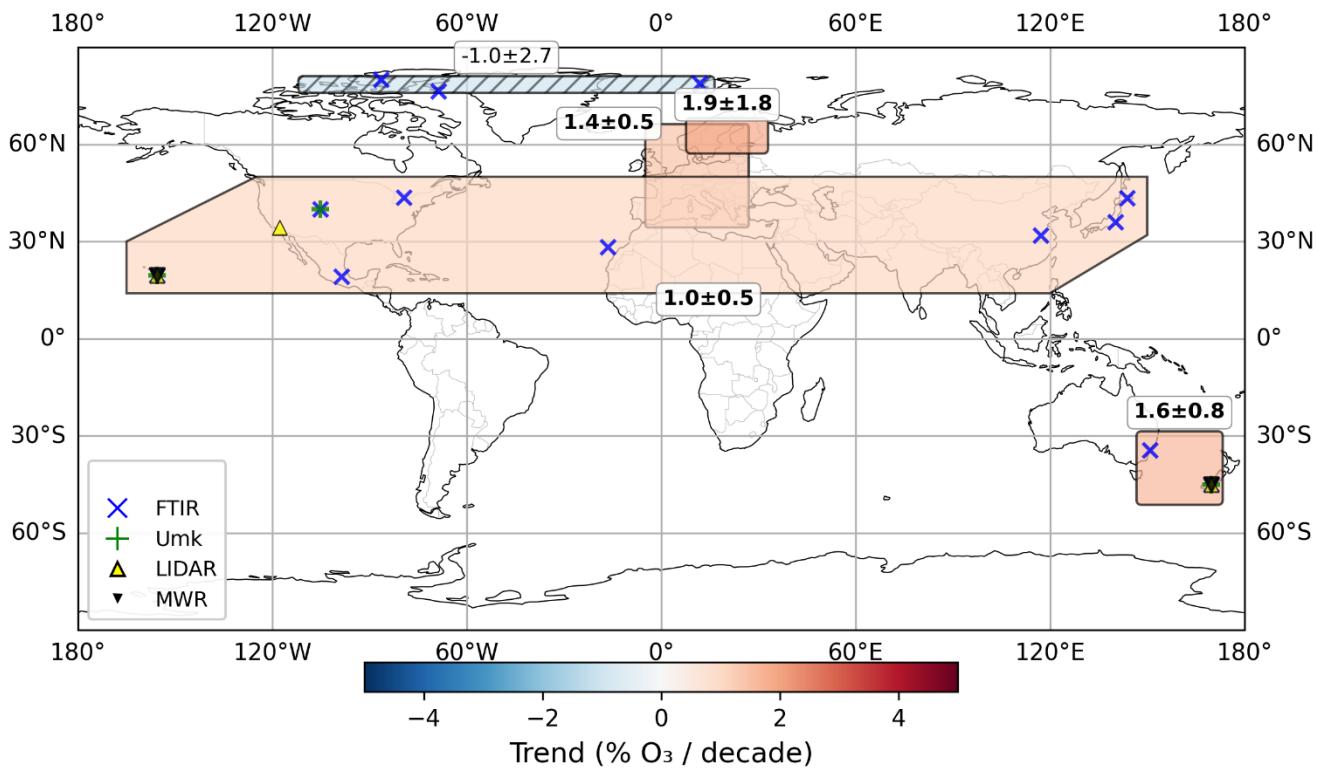



Figure 20: Ozone trends for the regional groups in the Upper Stratosphere (UpS). Bold numbers indicate significant trends.
690 The colour scale indicates the trend magnitude: red for positive and blue for negative values. Regions are hatched where trends are not significant at the 95% confidence level (2 sigma). The locations of the instruments merged into the composite are indicated by markers, except in Europe for readability.

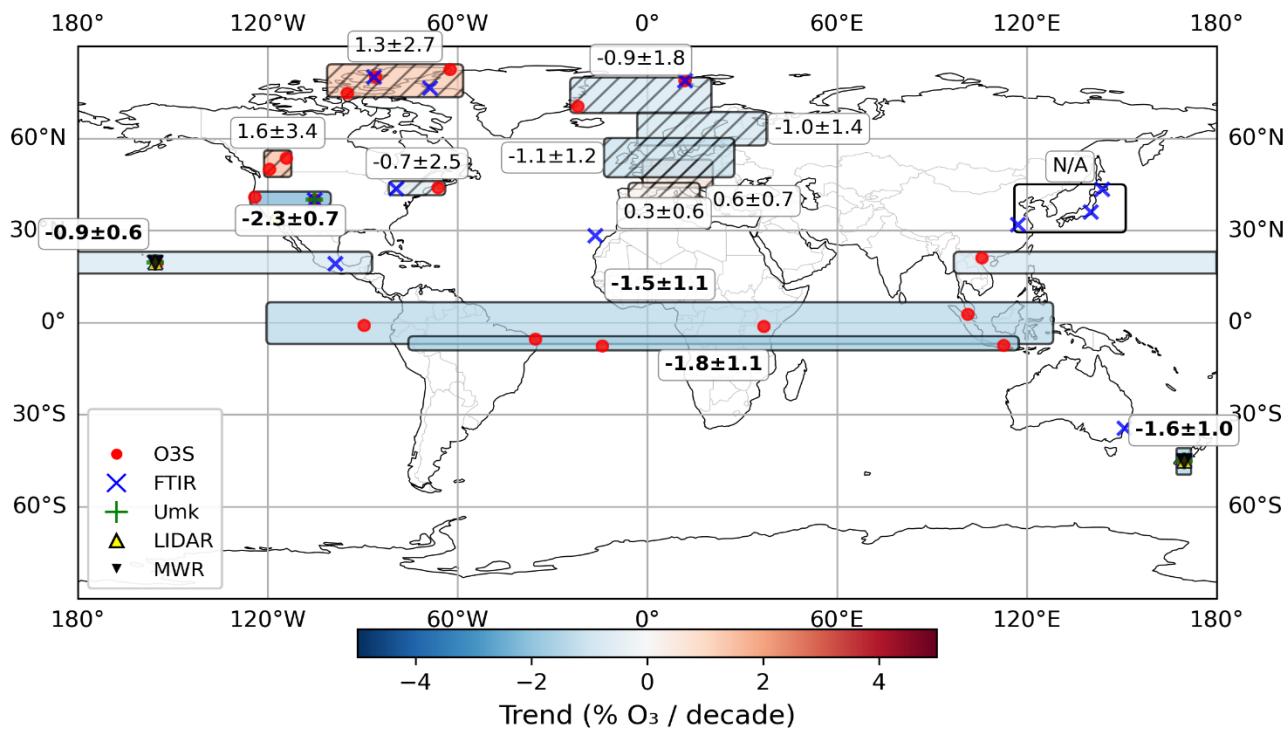



Figure 21: Same as figure 20 for layer MS

695

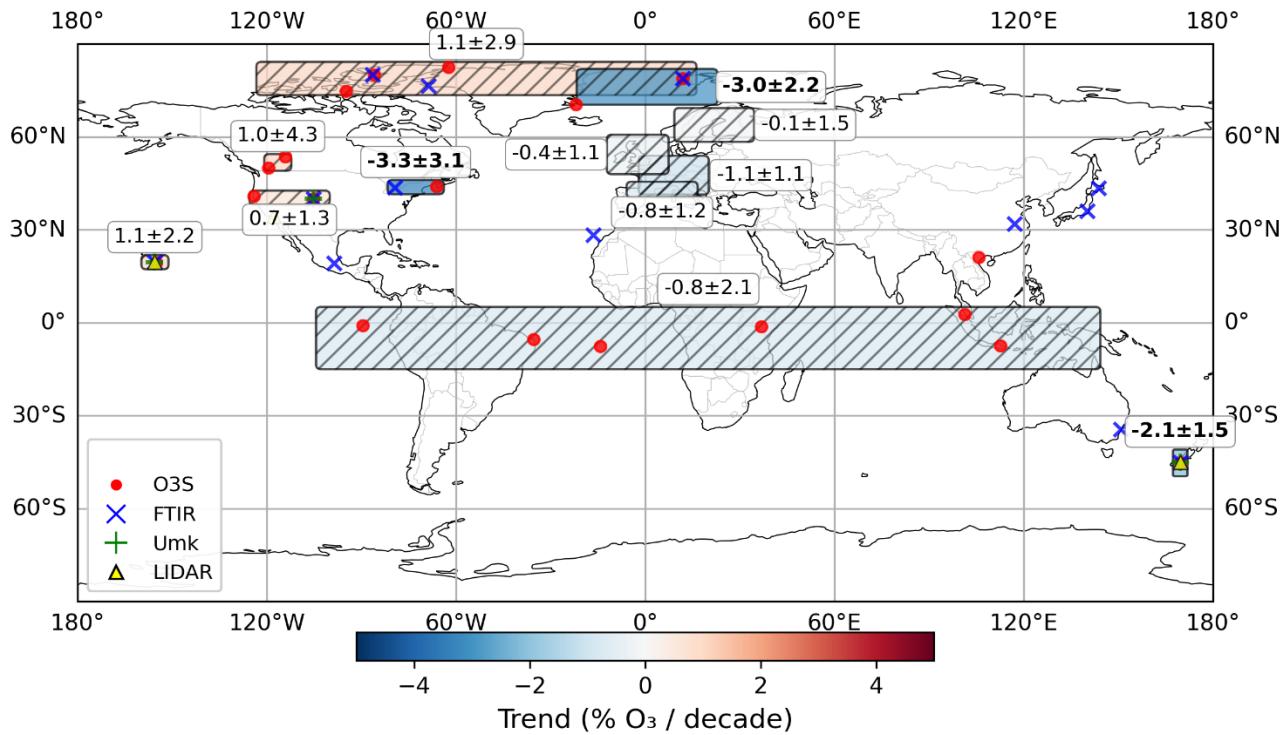



Figure 22: Same as figure 20 for layer LS

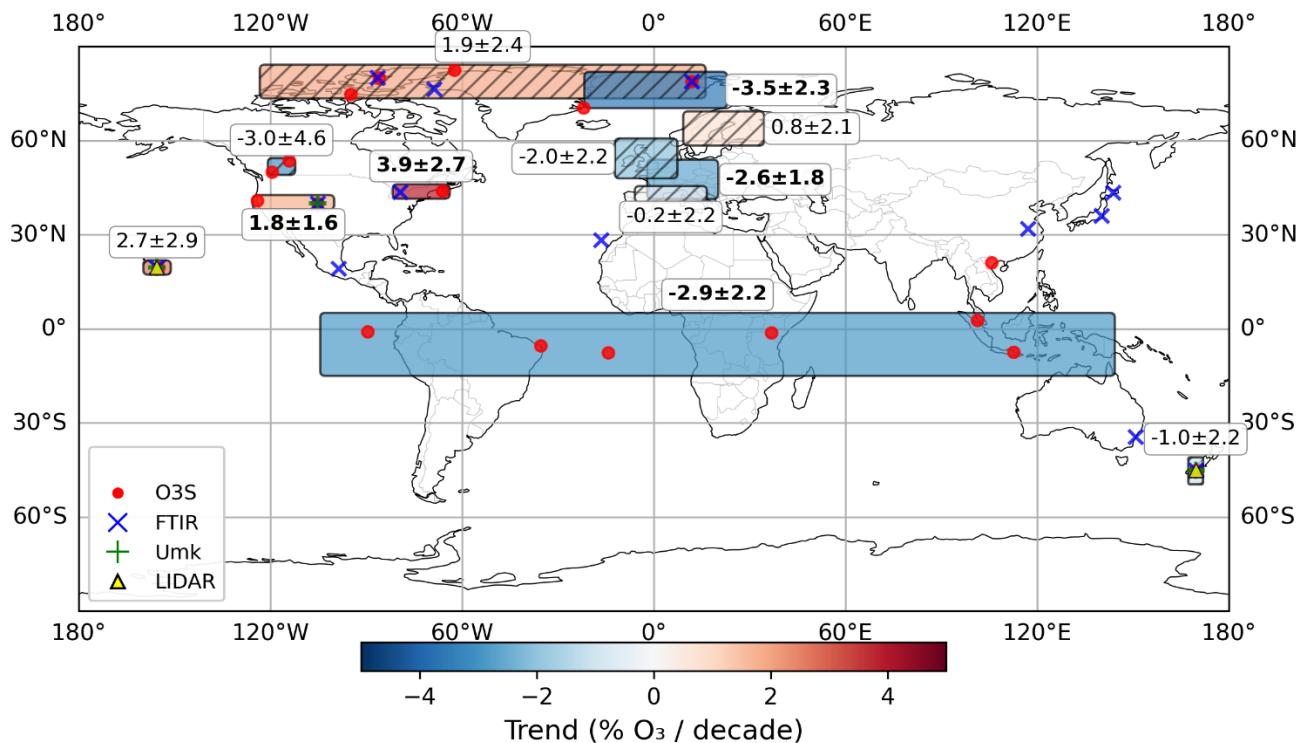



Figure 23: Same as figure 20 for layer UTLS

5 Conclusion

700   In this study, we use the BAyeSian Integrated and Consolidated (BASIC) merging method to build regional composites of ground-based ozone partial columns for the 2000–2024 period and compare its performance with the conventional weighted mean (WM). BASIC proves superior in handling the heterogeneity of ground-based networks, such as in the Central Europe LS case study, where BASIC's uncertainty dynamically adapts to the data consensus level. Using a Gaussian-mixture likelihood and PCA uncertainties for outliers detection, BASIC produces adaptive posterior uncertainties that can increase

705   monthly uncertainties when the instrumental consensus is poor, unlike the WM which does not adapt. While standard rejection techniques can easily remove obvious outliers, BASIC can also handle ambiguous values. Finally, despite larger composite uncertainties than the WM method, the trends uncertainties derived are smaller or comparable to those of the WM method, with an average reduction of 15.3 % in the BASIC trend uncertainty compared to WM's.

For the three groups of Lauder, Hawaii and Central Europe, we confirm a robust ozone recovery in the Upper Stratosphere

710   across all studied domains. In contrast, the Middle and Lower Stratosphere commonly show negative or mixed signals. In some cases, the combination of BASIC and Partial Columns definition is able to identify significance in trends when the trends derived for individual instrument time series are not significant. In other cases, significance results mainly from the noise reduction associated with the partial column definition. For example, both BASIC and the WM estimated a significant

negative trend in LS at Lauder, whereas satellite profile trends are not conclusive. In the tropics, Mauna Loa shows a
715   transition from positive UpS to negative MS ozone trends, with an inversion near 10 hPa. However, the UTLS signal
remains sensitive to the inclusion of the sparse FTIR dataset, the positive signal in TROPO and the choice of dynamical
proxies (NAO vs ENSO).

Global maps confirm the UpS recovery and show predominantly negative trends in the MS and LS. However, we find that
composite trends in the UTLS over North America are likely subject to a masking effect, whereby upper-tropospheric
720   positive trends hide a decrease in the lower-stratospheric. Our ground-based composite global maps compare well with the
satellite-composite global maps, although some limitations apply: the ground-based geographical coverage is uneven (with
major gaps in South America, Southern Africa, and Siberia), and winter data gaps at high latitudes due to polar nights inflate
trend uncertainties.

Overall, BASIC is a robust method to merge ground-based partial-column datasets for regional ozone trend detection,
725   providing an independent validation for global satellite assessments. However, BASIC is not able to reduce the uncertainties
when merging less than 3 instruments. This method is best suited to data-rich regions, and cannot be used as an alternative to
global coverage in regions where ground-based networks are absent.

Looking forward, this work highlights the potential of BASIC merging as a standard for heterogeneous ground networks.
Ground-based composites could then be compared to satellite-based composites. Future studies should focus on refining the
730   trend estimation techniques, for instance by applying Dynamical Linear Modeling (DLM). Additionally, simulation studies
that account for inter-instrument covariance would help to assess the representativeness of BASIC and WM composites.
Further studies will investigate the contribution of partial column trends to the total ozone column trends of merged
composites.

Appendix A: ENSO vs NAO at Hawaii

735   To assess the impact of the chosen dynamical proxy on trend estimates in the subtropical Pacific, we performed a sensitivity
analysis for the Hawaii group by replacing the North Atlantic Oscillation (NAO) proxy with the El Niño-Southern
Oscillation (ENSO) proxy, that goes with Aerosol Optical Depth (AOD).

For the original set oPC (Fig. A1), the positive trend magnitude in UpS is smaller when using ENSO compared to NAO,
though it remains statistically significant for both the BASIC and WM composites. The adjusted R2 values are comparable
740   between the two proxies for this layer. In MS, trends remain statistically significant and negative, with values similar to the
NAO case. The adjusted R2 is slightly higher with ENSO in MS, but the improvement is negligible (Fig. A3a). However, in
LS, the trend shifts to a non-significant negative value with a notably larger uncertainty envelope compared to the NAO
case. The R2 adjusted remains weak in this layer, consistently below 0.5 whatever the proxy. Finally, while the BASIC
composite still indicates a positive trend in the TROPO, it loses statistical significance when using ENSO.

745 For the alternative set aPC (Fig. A2), the use of ENSO almost doubles the uncertainty estimates in the aUpS compared to NAO, although the adjusted R2 values remain similar (Fig. A3b). Trends in the aMS do not change, but uncertainties are larger, with the same adjusted R2. The UTLS layer shows the most distinct sensitivity and provides key insights into discrepancies with the literature. While uncertainties are similar for both proxies in the composites, the adjusted R2 improves with ENSO, rising from approximately 0.6 to 0.7. Notably, individual instruments show significant divergence: the FTIR at

750 Mauna Loa in UpS shifts from a strong positive trend (+4 %/decade) with NAO to a negative trend (-4 %/decade) with ENSO. The Alternative Troposphere (a-TROPO) trend estimates are robust and show no significant changes between the two proxies.

Sofieva et al. (2025) reports negative trends in the tropical UTLS. Our primary analysis using NAO gave positive trends, raising the question of whether this divergence was driven by the proxy choice or dataset composition. The sensitivity

755 analysis shows that using the ENSO proxy shifts the trends to be more negative—aligning better with the sign reported in the literature—but does not render them statistically significant. Furthermore, the Mauna Loa FTIR record, which is included in our study but absent in Sofieva et al. (2025), exhibits extreme sensitivity to the proxy choice, flipping from positive with NAO to negative with ENSO. The changes in the R2 adjusted are little to none, depending on the partial column. Our findings suggests that the trends are highly sensitive to the proxies choice, and a combination of more proxies should be

760 done– outside the scope of this work. The discrepancy with satellite literature also results from the inclusion of the temporally sparse FTIR, introducing significant volatility.

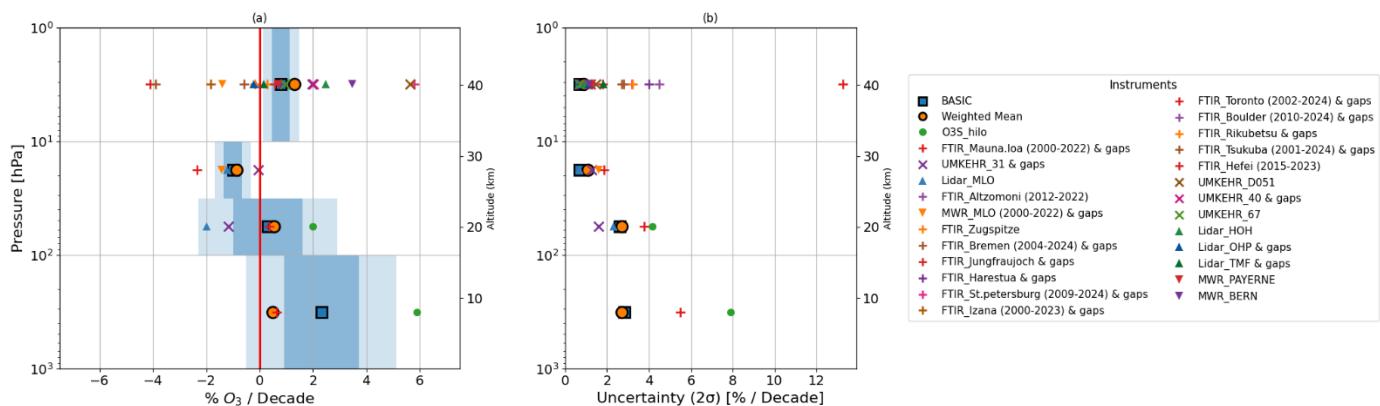



Figure A1: Trend estimates for the oPC of the regional group Hawaii, replacing the proxy NAO for ENSO. (a) 2000-2024

765 trends in % per decade. BASIC trend values are a blue square with black contour, with 1 sigma uncertainty in light blue shading and 2 sigma in darker blue shading. The WM trend is an orange circle with black contour. Trends for each individual dataset involved in the merged composite are plotted for comparison. (b) 2-sigma uncertainties of the composites and of each dataset are shown in % per decade. The legend highlights datasets with time ranges deviating from the 2000-2024 time period and/or substantial gaps.

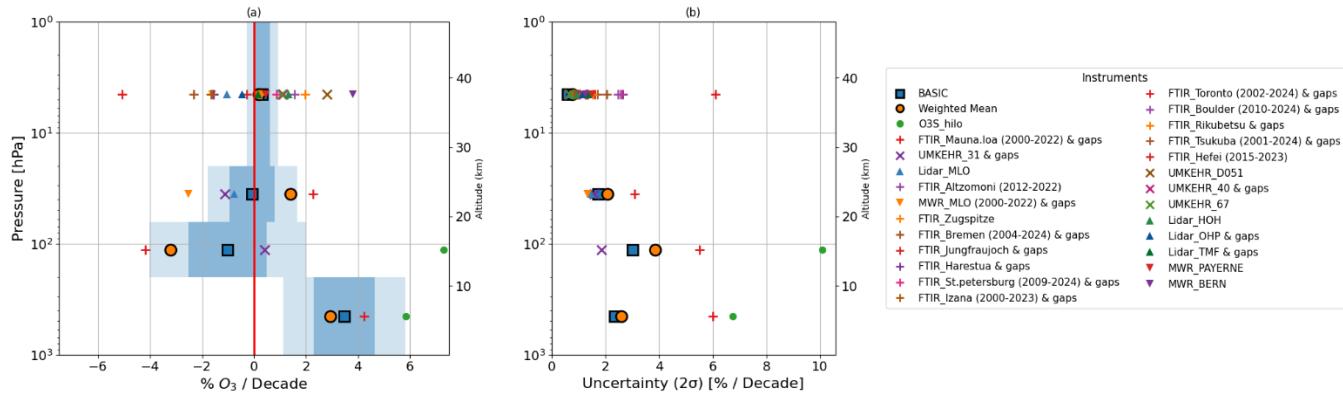



Figure A2: Same as A1 for the alternative set aPC

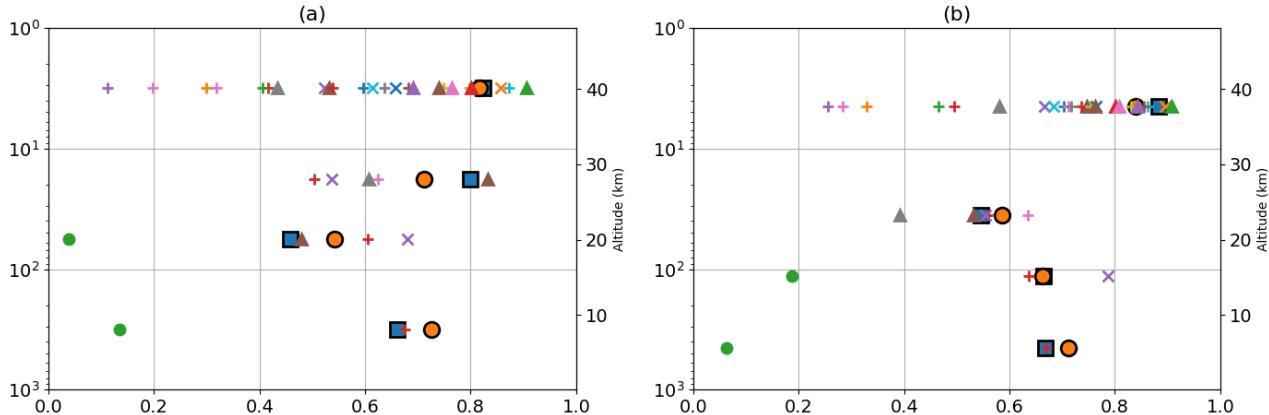



Figure A3: Adjusted R2 for BASIC, the weighted mean and the datasets of the Hawaii with ENSO, for (a) oPC and (b) aPC. Legends are similar to Figures A1 and A2.

Code, data, or code and data availability

All the datasets are available at LOTUS ftp website (contact the first author for access). At the stage of paper acceptance, all trend results will be provided in open-access repository. The code will also be available on GIT.

Supplement link

The link to the supplement will be included by Copernicus, if applicable.

Author contributions

CJ used CAMS to derive the correlations between the stations for all the partial columns. EMB computed the partial column time series for the MWR and for the Swiss Dobson Umkehr, CV and CJ for the FTIR, RVM for the ozonesondes, IP for the NOAA Dobson Umkehr and SGB, TL and WS for the Lidars. AV adapted to profiles the BASIC merging routine of Ball et al.. LM performed the regional BASIC merging and computed the trends. LM, EMB, CJ, CV, RVM, IP, TL and SGB all contributed to the writing of the paper. The other co-author are the PIs of the data records used in the study. All co-authors discussed the paper.

Competing interests

At least one of the (co-)authors is a member of the editorial board of Atmospheric Chemistry and Physics.

Acknowledgements

LM was funded by the Swiss National Science Foundation (SNSF) 200021L-228149. LM and GS are members of the Oeschger Center for Climate Change Research.

The FTIR monitoring program at Jungfraujoch has been primarily supported by the F.R.S. - FNRS (Brussels, Belgium), the GAW-CH program of MeteoSwiss (Switzerland) and the University of Liège. E.M. is a research director with F.R.S. - FNRS. The ULiège team thanks the International Foundation High Altitude Research Stations Jungfraujoch and Gornergrat (HFSJG, Bern, Switzerland) for supporting the facilities needed to perform the Fourier Transform InfraRed observations at Jungfraujoch. The lidar and sondes measurements at OHP are funded by CNRS (France). The National Center for Atmospheric Research is sponsored by the National Science Foundation. The NCAR FTS observation programs at Thule, GR, Boulder, CO and Mauna Loa, HI are supported under contract by the National Aeronautics and Space Administration (NASA). The Thule work is also supported by the NSF Office of Polar Programs (OPP). We wish to thank the Danish Meteorological Institute for support at the Thule site and NOAA for support of the MLO site. The Eureka FTIR measurements were made at the Polar Environment Atmospheric Research Laboratory (PEARL), primarily supported by ECCC, CSA, and NSERC. The Toronto FITR measurements were made at the University of Toronto Atmospheric Observatory, supported by NSERC and ECCC. Isamu Morino at NIES operates two FTIRs at the Tsukuba and Rikubetsu sites. The FTIR operations at the Tsukuba and Rikubetsu sites are supported in part by the GOSAT series project. We thank the team of PMOD/WRC for the measurement performed in Davos.

Several AI tools were used in this work, including chatGPT, Claude and Gemini, for example for sentence composition and literature research. Copilot pro was used within the code, to generate comments, write functions and help create some illustrations such as the maps. All the outputs were verified by the authors.

References

810

Arosio, C., Rozanov, A., Malinina, E., Weber, M., and Burrows, J. P.: Merging of ozone profiles from SCIAMACHY, OMPS and SAGE II observations to study stratospheric ozone changes, Atmos. Meas. Tech., 12, 2423–2444, https://doi.org/10.5194/amt-12-2423-2019, 2019.

815 Ball, W. T., Alsing, J., Mortlock, D. J., Rozanov, E. V., Tummon, F., and Haigh, J. D.: Reconciling differences in stratospheric ozone composites, Atmos. Chem. Phys., 17, 12269–12302, https://doi.org/10.5194/acp-17-12269-2017, 2017

Ball, W. T., Rozanov, E., Alsing, J. A., Marsh, D. R., Tummon, F., Mortlock, D. J., et al. (2019). The upper stratospheric solar cycle ozone response. Geophysical Research Letters, 46, 1831–1841. https://doi.org/10.1029/2018GL081501

820

Ball, W. T., Chiodo, G., Abalos, M., Alsing, J., and Stenke, A.: Inconsistencies between chemistry–climate models and observed lower stratospheric ozone trends since 1998, Atmos. Chem. Phys., 20, 9737–9752, https://doi.org/10.5194/acp-20-9737-2020, 2020.

825 Björklund, R., Vigouroux, C., Effertz, P., García, O. E., Geddes, A., Hannigan, J., Miyagawa, K., Kotkamp, M., Langerock, B., Nedoluha, G., Ortega, I., Petropavlovskikh, I., Poyraz, D., Querel, R., Robinson, J., Shiona, H., Smale, D., Smale, P., Van Malderen, R., and De Mazière, M.: Intercomparison of long-term ground-based measurements of total, tropospheric, and stratospheric ozone at Lauder, New Zealand, Atmos. Meas. Tech., 17, 6819–6849, https://doi.org/10.5194/amt-17-6819-2024, 2024.

830

Davis, S. M., Rosenlof, K. H., Hassler, B., Hurst, D. F., Read, W. G., Vömel, H., Selkirk, H., Fujiwara, M., and Damadeo, R.: The Stratospheric Water and Ozone Satellite Homogenized (SWOOSH) database: a long-term database for climate studies, Earth Syst. Sci. Data, 8, 461–490, https://doi.org/10.5194/essd-8-461-2016, 2016.

835 Dietmüller, S., Garny, H., Eichinger, R., and Ball, W. T.: Analysis of recent lower-stratospheric ozone trends in chemistry climate models, Atmos. Chem. Phys., 21, 6811–6837, https://doi.org/10.5194/acp-21-6811-2021, 2021.

Emili, E., Barret, B., Massart, S., Le Flochmoen, E., Piacentini, A., El Amraoui, L., Pannekoucke, O., and Cariolle, D.: Combined assimilation of IASI and MLS observations to constrain tropospheric and stratospheric ozone in a global chemical 840 transport model, Atmos. Chem. Phys., 14, 177–198, https://doi.org/10.5194/acp-14-177-2014, 2014.

Stephanie Evan et al., Rapid ozone depletion after humidification of the stratosphere by the Hunga Tonga Eruption.Science 382, eadg2551(2023). DOI:10.1126/science.adg2551

Frith, S. M., Bhartia, P. K., Oman, L. D., Kramarova, N. A., McPeters, R. D., and Labow, G. J.: Model-based climatology of diurnal variability in stratospheric ozone as a data analysis tool, Atmos. Meas. Tech., 13, 2733–2749, 845 https://doi.org/10.5194/amt-13-2733-2020, 2020.

Froidevaux, L., Anderson, J., Wang, H.-J., Fuller, R. A., Schwartz, M. J., Santee, M. L., Livesey, N. J., Pumphrey, H. C., Bernath, P. F., Russell III, J. M., and McCormick, M. P.: Global OZone Chemistry And Related trace gas Data records for the Stratosphere (GOZCARDS): methodology and sample results with a focus on HCl, H2O, and O3, Atmos. Chem. Phys., 850 15, 10471–10507, https://doi.org/10.5194/acp-15-10471-2015, 2015.

Godin-Beekmann, S., Azouz, N., Sofieva, V. F., Hubert, D., Petropavlovskikh, I., Effertz, P., Ancellet, G., Degenstein, D. A., Zawada, D., Froidevaux, L., Frith, S., Wild, J., Davis, S., Steinbrecht, W., Leblanc, T., Querel, R., Tourpali, K., Damadeo, R., Maillard Barras, E., Stübi, R., Vigouroux, C., Arosio, C., Nedoluha, G., Boyd, I., Van Malderen, R., Mahieu, 855 E., Smale, D., and Sussmann, R.: Updated trends of the stratospheric ozone vertical distribution in the 60° S–60° N latitude range based on the LOTUS regression model , Atmos. Chem. Phys., 22, 11657–11673, https://doi.org/10.5194/acp-22-11657-2022, 2022.

Godin-Beekmann, S., Porteneuve, J., and Garnier, A.: Systematic DIAL ozone measurements at Observatoire de Haute-860 Provence, J. Env. Monitor., 5, 57–67, 2003

Gordon, I., Rothman, L., Hargreaves, R., Hashemi, R., Karlovets, E., Skinner, F., Conway, E., Hill, C., Kochanov, R., Tan, Y., Wcisło, P., Finenko, A., Nelson, K., Bernath, P., Birk, M., Boudon, V., Campargue, A., Chance, K., Coustenis, A., Drouin, B., Flaud, J., Gamache, R., Hodges, J., Jacquemart, D., Mlawer, E., Nikitin, A., Perevalov, V., Rotger, M., 865 Tennyson, J., Toon, G., Tran, H., Tyuterev, V., Adkins, E., Baker, A., Barbe, A., Canè, E., Császár, A., Dudaryonok, A., Egorov, O., Fleisher, A., Fleurbaey, H., Foltynowicz, A., Furtenbacher, T., Harrison, J., Hartmann, J., Horneman, V., Huang, X., Karman, T., Karns, J., Kassi, S., Kleiner, I., Kofman, V., Kwabia-Tchana, F., Lavrentieva, N., Lee, T., Long, D., Lukashevskaya, A., Lyulin, O., Makhnev, V., Matt, W., Massie, S., Melosso, M., Mikhailenko, S., Mondelain, D., Müller, H., Naumenko, O., Perrin, A., Polyansky, O., Raddaoui, E., Raston, P., Reed, Z., Rey, M., Richard, C., Tóbiás, R., Sadiek, I., 870 Schwenke, D., Starikova, E., Sung, K., Tamassia, F., Tashkun, S., Vander Auwera, J., Vasilenko, I., Vigasin, A., Villanueva, G., Vispoel, B., Wagner, G., Yachmenev, A., and Yurchenko, S.: The HITRAN2020 molecular spectroscopic database, J. Quant. Spectrosc. Ra., 277, 107949, https://doi.org/10.1016/j.jqsrt.2021.107949, 2022.

Inness A. et al, 2019, https://doi.org/10.5194/acp-19-3515-2019. Footnote necessary: Copernicus Atmosphere Monitoring 875 Service (2020): CAMS global reanalysis (EAC4) monthly averaged fields. Copernicus Atmosphere Monitoring Service (CAMS) Atmosphere Data Store, DOI: 10.24381/fd75fff2 (Accessed on 14-03-2025).

Jonas, C., Vigouroux, C., Langerock, B., Björklund, R., Boynard, A., Carlund, T., De Mazière, M., Effertz, P., Errera, Q., Frey, M., Hannigan, J. W., Jepsen, N., Kivi, R., Lyall, N., Palm, M., Prignon, M., Sofieva, V. F., Strong, K., Svendby, T., Tarasick, D., Thölix, L., Van Malderen, R., Virolainen, Y., von Löwis, S., and Zhao, X.: Detection of ozone recovery in the
880   Arctic from ground-based measurements, EGUsphere [preprint], https://doi.org/10.5194/egusphere-2025-6473, 2026.

Keppens, A., Hubert, D., Granville, J., Nath, O., Lambert, J.-C., Wespes, C., Coheur, P.-F., Clerbaux, C., Boynard, A., Siddans, R., Latter, B., Kerridge, B., Di Pede, S., Veefkind, P., Cuesta, J., Dufour, G., Heue, K.-P., Coldewey-Egbers, M., Loyola, D., Orfanoz-Cheuquelaf, A., Maratt Satheesan, S., Eichmann, K.-U., Rozanov, A., Sofieva, V. F., Ziemke, J. R.,
885   Inness, A., Van Malderen, R., and Hoffmann, L.: Harmonisation of sixteen tropospheric ozone satellite data records, Atmos. Meas. Tech., 18, 6893–6916, https://doi.org/10.5194/amt-18-6893-2025, 2025.

Kramarova, N. A., Frith, S. M., Bhartia, P. K., McPeters, R. D., Taylor, S. L., Fisher, B. L., Labow, G. J., and DeLand, M. T.: Validation of ozone monthly zonal mean profiles obtained from the version 8.6 Solar Backscatter Ultraviolet algorithm,
890   Atmos. Chem. Phys., 13, 6887–6905, https://doi.org/10.5194/acp-13-6887-2013, 2013.

Leblanc, T., Sica, R. J., van Gijsel, J. A. E., Godin-Beekmann, S., Haefele, A., Trickl, T., Payen, G., and Gabarrot, F.: Proposed standardized definitions for vertical resolution and uncertainty in the NDACC lidar ozone and temperature algorithms – Part 1: Vertical resolution, Atmos. Meas. Tech., 9, 4029–4049, https://doi.org/10.5194/amt-9-4029-2016, 2016.
895

Maillard Barras, E., Haefele, A., Nguyen, L., Tummon, F., Ball, W. T., Rozanov, E. V., Rüfenacht, R., Hocke, K., Bernet, L., Kämpfer, N., Nedoluha, G., and Boyd, I.: Study of the dependence of long-term stratospheric ozone trends on local solar time, Atmos. Chem. Phys., 20, 8453–8471, https://doi.org/10.5194/acp-20-8453-2020, 2020.

900   McPeters, R. D., and G. J. Labow (2012), Climatology 2011: An MLS and sonde derived ozone climatology for satellite retrieval algorithms, J. Geophys. Res., 117, D10303, doi:10.1029/2011JD017006.

Millán, L., Hoor, P., Hegglin, M. I., Manney, G. L., Jeffery, P. S., Weyland, F. M., et al. (2025). Ozone trends in the upper troposphere-lower stratosphere using equivalent latitude-potential temperature coordinates. Geophysical Research Letters,
905   52, e2025GL118651. https://doi.org/10.1029/2025GL118651

NDACC   observation   capabilities,   https://ndacc.larc.nasa.gov/data/observational-capabilities, https://ndacc.larc.nasa.gov/sites/default/files/docs/data/NDACC_ObsCap_20201108_small.pdf

910   Petropavlovskikh, I., P. K. Bhartia, and J. DeLuisi (2005), New Umkehr ozone profile retrieval algorithm optimized for climatological studies, Geophys. Res. Lett., 32, L16808, doi:10.1029/2005GL023323

Petropavlovskikh, I, Evans R. D., Carbaugh G. L., Maillard E. And Stubi R., 2008, Towards a better knowledge of Umkehr measurements: A detailed study of data from thirteen Dobson intercomparisons, World Meteorological Organization Global
915   Atmosphere Watch, GAW No. 180.

Petropavlovskikh, Irina & Evans, Robert & Mcconville, Glen & Miyagawa, Koji & Oltmans, Samuel. (2009). Effect of the out-of-band stray light on the retrieval of the Umkehr Dobson ozone profiles. International Journal of Remote Sensing - INT J REMOTE SENS. 30. 6461-6482. 10.1080/01431160902865806.

920
Petropavlovskikh, I., Evans, R., McConville, G., Oltmans, S., Quincy, D., Lantz, K., Disterhoft, P., Stanek, M., and Flynn, L.: Sensitivity of Dobson and Brewer Umkehr ozone profile retrievals to ozone cross-sections and stray light effects, Atmos. Meas. Tech., 4, 1841–1853, https://doi.org/10.5194/amt-4-1841-2011, 2011.

925   Petropavlovskikh, I., Wild, J. D., Abromitis, K., Effertz, P., Miyagawa, K., Flynn, L. E., Maillard Barras, E., Damadeo, R., McConville, G., Johnson, B., Cullis, P., Godin-Beekmann, S., Ancellet, G., Querel, R., Van Malderen, R., and Zawada, D.: Ozone trends in homogenized Umkehr, ozonesonde, and COH overpass records, Atmos. Chem. Phys., 25, 2895–2936, https://doi.org/10.5194/acp-25-2895-2025, 2025.

930   Rodgers, C. D.: Inverse Methods for Atmospheric Sounding – Theory and Practice, vol. 2 of Series on Atmospheric Oceanic and Planetary Physics, World Scientific Publishing Co. Pte. Ltd., Singapore, https://doi.org/10.1142/9789812813718, 2000.

Rigby, M., Park, S., Saito, T. et al. Increase in CFC-11 emissions from eastern China based on atmospheric observations. Nature 569, 546–550 (2019). https://doi.org/10.1038/s41586-019-1193-4
935
Salawitch, R. J., and Coauthors, 2025: The Imminent Data Desert: The Future of Stratospheric Monitoring in a Rapidly Changing World. Bull. Amer. Meteor. Soc., 106, E540–E563, https://doi.org/10.1175/BAMS-D-23-0281.1.

Sauvageat, Eric & Albers, Roland & Kotiranta, Mikko & Hocke, Klemens & Gómez, Michael & Nedoluha, Gerald & Murk,
940   Axel. (2021). Comparison of Three High Resolution Real-Time Spectrometers for Microwave Ozone Profiling Instruments. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing. PP. 1-1. 10.1109/JSTARS.2021.3114446.

945 Sauvageat, E., Maillard Barras, E., Hocke, K., Haefele, A., and Murk, A.: Harmonized retrieval of middle atmospheric ozone from two microwave radiometers in Switzerland, Atmos. Meas. Tech., 15, 6395–6417, https://doi.org/10.5194/amt-15-6395-2022, 2022.

Sauvageat, E., Hocke, K., Maillard Barras, E., Hou, S., Errera, Q., Haefele, A., and Murk, A.: Microwave radiometer
950 observations of the ozone diurnal cycle and its short-term variability over Switzerland, Atmos. Chem. Phys., 23, 7321–7345, https://doi.org/10.5194/acp-23-7321-2023, 2023.

Schranz, F., Fernandez, S., Kämpfer, N., and Palm, M.: Diurnal variation in middle-atmospheric ozone observed by ground-based microwave radiometry at Ny-Ålesund over 1 year, Atmos. Chem. Phys., 18, 4113–4130, https://doi.org/10.5194/acp-
955 18-4113-2018, 2018.

Smit, H. G. J., Thompson, A. M., and the ASOPOS 2.0 Panel: Ozonesonde Measurement Principles and Best Operational Practices, WMO Global Atmosphere Watch Report Series, No. 268, World Meteorological Organization, Geneva, https://library.wmo.int/idurl/4/57720 (last access: 19 November 2025), 2021.
960
Smit, H. G. J., Poyraz, D., Van Malderen, R., Thompson, A. M., Tarasick, D. W., Stauffer, R. M., Johnson, B. J., and Kollonige, D. E.: New insights from the Jülich Ozone Sonde Intercomparison Experiment: calibration functions traceable to one ozone reference instrument, Atmos. Meas. Tech., 17, 73–112, https://doi.org/10.5194/amt-17-73-2024, 2024.

965 Sofieva, V. F., Szeląg, M., Tamminen, J., Kyrölä, E., Degenstein, D., Roth, C., Zawada, D., Rozanov, A., Arosio, C., Burrows, J. P., Weber, M., Laeng, A., Stiller, G. P., von Clarmann, T., Froidevaux, L., Livesey, N., van Roozendael, M., and Retscher, C.: Measurement report: regional trends of stratospheric ozone evaluated using the MErged GRIdded Dataset of Ozone Profiles (MEGRIDOP), Atmos. Chem. Phys., 21, 6707–6720, https://doi.org/10.5194/acp-21-6707-2021, 2021.

970 Sofieva, V. F., Szelag, M. E., Kramarova, N., Damadeo, R., Steinbrecht, W., Petropavlovskikh, I., Vigouroux, C., Maillard Barras, E., Zawada, D., Tourpali, K., Frith, S. M., Wild, J. D., Davis, S. M., Arosio, C., Weber, M., Rozanov, A., Auffarth, B., Froidevaux, L., Fuller, R., Degenstein, D., Dube, K., Effertz, P., Leblanc, T., Ancellet, G., Godin-Beekmann, S., McConville, G., Querel, R., Smale, D., DeBacker, M.-R., Mahieu, E., and Sussmann, R.: Updated global and regional trends of stratospheric ozone profiles, EGUsphere [preprint], https://doi.org/10.5194/egusphere-2025-5963, 2025.
975

Szeląg, M. E., Sofieva, V. F., Degenstein, D., Roth, C., Davis, S., and Froidevaux, L.: Seasonal stratospheric ozone trends over 2000–2018 derived from several merged data sets, Atmos. Chem. Phys., 20, 7035–7047, https://doi.org/10.5194/acp-20-7035-2020, 2020.

980 Thompson, A. M., Stauffer, R. M., Kollonige, D. E., Ziemke, J. R., Johnson, B. J., Morris, G. A., Cullis, P., Cazorla, M., Diaz, J. A., Piters, A., Nedeljkovic, I., Warsodikromo, T., Raimundo Silva, F., Northam, E. T., Benjamin, P., Mkololo, T., Machinini, T., Félix, C., Romanens, G., Nyadida, S., Brioude, J., Evan, S., Metzger, J.-M., Dindang, A., Mahat, Y. B., Sammathuria, M. K., Zakaria, N. B., Komala, N., Ogino, S.-Y., Quyen, N. T., Mani, F. S., Vuiyasawa, M., Nardini, D., Martinsen, M., Kuniyuki, D. T., Müller, K., Wolff, P., and Sauvage, B.: Tropical tropospheric ozone trends (1998 to 2023):
985 new perspectives from SHADOZ, IAGOS and OMI/MLS observations, Atmos. Chem. Phys., 25, 18475–18507, https://doi.org/10.5194/acp-25-18475-2025, 2025.

Van Malderen, R., Thompson, A. M., Kollonige, D. E., Stauffer, R. M., Smit, H. G. J., Maillard Barras, E., Vigouroux, C., Petropavlovskikh, I., Leblanc, T., Thouret, V., Wolff, P., Effertz, P., Tarasick, D. W., Poyraz, D., Ancellet, G., De Backer,
990 M.-R., Evan, S., Flood, V., Frey, M. M., Hannigan, J. W., Hernandez, J. L., Iarlori, M., Johnson, B. J., Jones, N., Kivi, R., Mahieu, E., McConville, G., Müller, K., Nagahama, T., Notholt, J., Piters, A., Prats, N., Querel, R., Smale, D., Steinbrecht, W., Strong, K., and Sussmann, R.: Global ground-based tropospheric ozone measurements: reference data and individual site trends (2000–2022) from the TOAR-II/HEGIFTOM project, Atmos. Chem. Phys., 25, 7187–7225, https://doi.org/10.5194/acp-25-7187-2025, 2025a.
995

Van Malderen, R., Zang, Z., Chang, K.-L., Björklund, R., Cooper, O. R., Liu, J., Maillard Barras, E., Vigouroux, C., Petropavlovskikh, I., Leblanc, T., Thouret, V., Wolff, P., Effertz, P., Gaudel, A., Tarasick, D. W., Smit, H. G. J., Thompson, A. M., Stauffer, R. M., Kollonige, D. E., Poyraz, D., Ancellet, G., De Backer, M.-R., Frey, M. M., Hannigan, J. W., Hernandez, J. L., Johnson, B. J., Jones, N., Kivi, R., Mahieu, E., Morino, I., McConville, G., Müller, K., Murata, I., Notholt,
1000 J., Piters, A., Prignon, M., Querel, R., Rizi, V., Smale, D., Steinbrecht, W., Strong, K., and Sussmann, R.: Ground-based tropospheric ozone measurements: regional tropospheric ozone column trends from the TOAR-II/HEGIFTOM homogenized datasets, Atmos. Chem. Phys., 25, 9905–9935, https://doi.org/10.5194/acp-25-9905-2025 , 2025b.

Vigouroux, C., Blumenstock, T., Coffey, M., Errera, Q., García, O., Jones, N. B., Hannigan, J. W., Hase, F., Liley, B.,
1005 Mahieu, E., Mellqvist, J., Notholt, J., Palm, M., Persson, G., Schneider, M., Servais, C., Smale, D., Thölix, L., and De Mazière, M.: Trends of ozone total columns and vertical distribution from FTIR observations at eight NDACC stations around the globe, Atmos. Chem. Phys., 15, 2915–2933, https://doi.org/10.5194/acp-15-2915-2015, 2015.

Wang, J., Chiodo, G., Ayarzagüena, B., Ball, W. T., Diallo, M., Hassler, B., Keeble, J., Nowack, P., Orbe, C., Vattioni, S., and Sukhodolov, T.: Exploring ozone–climate interactions in idealized CMIP6 DECK experiments, Atmos. Chem. Phys., 25,

1010   17819–17844, https://doi.org/10.5194/acp-25-17819-2025, 2025.

Weatherhead, E. C., Bodeker, G. E., Fassò, A., Chang, K.-L., Lazo, J. K., Clack, C. T. M., Hurst, D. F., Hassler, B., English, J. M., and Yorgun, S.: Spatial coverage of monitoring networks: A climate observing system simulation experiment, J. Appl. Meteorol. Clim., 56, 3211–3228, https://doi.org/10.1175/JAMC-D-17-0040.1, 2017.

1015   Wilks, D. S. (2019). Statistical Methods in the Atmospheric Sciences (4th ed.). Elsevier.