Ozone trends in homogenized Umkehr, Ozonesonde, and COH overpass records

Irina Petropavlovskikh[1,2], Jeannette D. Wild[3,4], Kari Abromitis[1,2], Peter Effertz[1,2], Koji Miyagawa[5], Lawrence E. Flynn[4], Eliane Maillard Barras[6], Robert Damadeo[7], Glen McConville[1,2], Bryan Johnson[2], Patrick Cullis[2], Sophie Godin-Beckmann[8], Gerard Ancellet[8], Richard Querel[9], Roeland Van Malderen[10], Daniel Zawada[11]

[1] CIRES, University of Colorado, Boulder, CO, USA
[2]NOAA, Global Monitoring Lab, Boulder, CO, USA
[3] University of Maryland, Earth System Science Interdisciplinary Center (ESSIC/CISESS), College Park, MD, USA
[4]NOAA/NESDIS/Center for Satellite Applications and Research (STAR), College Park, MD, USA
[5]National Institute for Environmental Studies (**NIES**), Tsukuba, Japan
[6] Federal Office of Meteorology and Climatology, MeteoSwiss, Payern, Switzerland
[7] NASA Langley Research Center, Hampton, VA, USA
[8] LATMOS Sorbonne Université, UVSQ-CNRS/INSU, Paris, France.
[9] National Institute of Water & Atmospheric Research (NIWA), Lauder, New Zealand
[10]Royal Meteorological Institute of Belgium, Uccle (Brussels), Belgium
[11]University of Saskatchewan, Saskatoon, SK, Canada

Correspondence to: Irina Petropavlovskikh (Irina.petro@Noaa.gov)

Abstract. This study presents an updated evaluation of stratospheric ozone profile trends at Arosa/Davos/Hohenpeißenberg, Switzerland/Germany, Observatory de Haute Provence (OHP), France, Boulder, Colorado, Mauna Loa Observatory (MLO) and Hilo, Hawaii, and Lauder, New Zealand with focus on the ozone recovery period post 2000. Trends are derived using vertical ozone profiles from NOAA's Dobson Network via the Umkehr method (with a recent new homogenization), ozonesondes, and the NOAA COHesive SBUV/OMPS satellite-based record (COH) sampled to match geographical coordinates of the ground-based stations used in this study. Analyses of long-term changes in stratospheric ozone time series were performed using the updated version (0.8.0) of the Long-term Ozone Trends and Uncertainties in the Stratosphere (LOTUS) Independent Linear Trend (ILT) regression model. This study finds a consistency of the trends derived from the different observational records, which is a key factor to the understanding of the recovery of the ozone layer after the implementation of the Montreal Protocol and its amendments that control ozone-depleting substances production and release into the atmosphere. The Northern Hemispheric Umkehr records of Arosa/Davos, OHP, and MLO all show positive trends in the mid to upper stratosphere with trends peaking at ~+2%/decade. Although the upper stratospheric ozone trends derived from COH satellite records are more positive than those detected by the Umkehr system, the agreement is within the two times standard error uncertainty. Umkehr trends in the upper stratosphere at Boulder and Lauder are positive but not statistically significant, while COH trends are larger and statistically significant (within 2 times standard error uncertainty). In the lower stratosphere, trends derived from Umkehr and ozonesonde records are mostly negative (except for positive ozonesonde trends at OHP), however the uncertainties are quite large. Additional dynamical proxies were investigated in the LOTUS model at five ground-based sites. The use of additional proxies did not significantly change trends, but equivalent latitude reduced the uncertainty of the Umkehr and COH trends in the

Deleted: -

Deleted: University of Maryland, Earth System Science Interdisciplinary Center (ESSIC), College Park, MD, USA¶

Deleted: NEIS, Tsukuba-City, Japan

Formatted: Font: Times New Roman

Deleted:

Deleted: sigma

Deleted: sigma

Deleted:

upper stratosphere and at higher latitudes. In lower layers, additional predictors (tropopause pressure for all stations, two extra components of Quasi-Biennial Oscillation at MLO, Arctic Oscillation at Arosa/Davos, OHP and MLO) improve the model fit and reduce trend uncertainties as seen by Umkehr and sonde.

1 Introduction

The WMO Ozone Assessments (WMO, 2018; WMO, 2022), indicate that for some geographical regions, the stratospheric ozone layer is recovering in accordance with the reduction of ozone depleting substances (ODS) whose production was restricted by the Montreal Protocol and its amendments. The US Clean Air Act requires NOAA to monitor prohibited chemicals and the ozone layer to ensure the success of the Montreal Protocol. NOAA's long-term network of measurements helps to interpret total column and vertically resolved ozone changes and link ozone recovery to the reduction of ODS levels in the stratosphere, changes in the lower stratosphere that are associated with climate changes, and to the increases in the troposphere that are influenced by the stratosphere/troposphere exchange and long-range transported pollution. The ongoing recovery of the stratospheric ozone layer is of great importance to human health (i.e. cancer from enhanced UV exposures, Madronich et al., 2021), the sustained production of crops, and the success of fisheries (dangerous algae blooms). For more information see the Environmental Effects Assessment Panel 2022 Quadrennial Assessment (EEAP, 2023).

The Long-term Ozone Trends and Uncertainties in the Stratosphere (LOTUS) study was initiated under Stratosphere-troposphere Processes And their Role in Climate (SPARC, changed to APARC in 2024) project to reconcile the differences in defining trend uncertainties between methods outlined in the WMO Assessment (WMO, 2014) and the SPARC/IO3C/IGACO-O3/NDACC (SI2N) study (Harris et al., 2015). Phase 1 focused on developing best practices for data merging, trend determination and error analyses. Results focused on analysis of broad latitudinal regions, near global, Northern and Southern Hemisphere, and Tropics as were used in the SI2N studies. Results are found in the 2019 report (SPARC/IO3C/GAW, 2019). Phase 2 refined the trend models, and extended the study to gridded, and GB ozone data sets. The development of methods used in trend detection is built on the community knowledge gained during the Tiger Team project in early 1990s (Reinsel et al., 2005), collaborations through the SPARC, WMO and IO3C supported LOTUS activity (Hassler et al., 2014; Harris et al., 2015; Godin-Beekmann et al., 2022) and the most recent contributions to the WMO Ozone assessment analyses published in Chapter 3, "Update on Global Ozone: Past, Present and Future" (Hassler et al., 2022).

Deleted: Studies of ozone recovery require long-term datasets often consisting of data merged from several instruments, or from a single instrument type on multiple satellite platforms, or at a ground-based (GB) station. 2011 saw the initiation of the SPARC/IO3C/IGACO-O3/NDACC (SI2N) activity to evaluate ozone trends in the depletion and recovery phases using both GB and merged satellite observations. The resulting report (Harris et al., 2015) emerged near the release of the 2014 WMO Ozone Assessment (WMO, 2014). The two studies resulted in broadly similar trend values in both the depletion and recovery phases. But the WMO report determined the recovery trends to be statistically significant, whereas the SI2N study did not. The discrepancy centered on differing error analysis techniques for the merged datasets: SI2N using distribution of the individual variances around the mean and WMO using weighted mean of the individual standard deviations.¶

Deleted: sought

Deleted: SI2N

Understanding the causes of the differences between GB and satellite records can create improvements not only in the internal consistency of data sets, but also in the uncertainties of overall ozone trends. Further, development of techniques to directly assess uncertainties in the merged records resulting from discrepancies that cannot be completely reconciled, such as small relative drifts and differences resulting from coordinate transformations and sampling differences, allows for a more precise estimate of significance of the mean trends. For the GB and satellite data used in the 2019 LOTUS Report, information on stability and drifts of the measurement was incomplete. The homogenization of many ozonesonde records was recently addressed and data were reprocessed (Tarasick et al., 2016; Van Malderen et al., 2016; Witte et al., 2017; Sterling et al., 2018; Witte et al., 2018; Ancellet et al., 2022) while some instrumental artifacts still need to be addressed (Smit, 2021).

The first attempt to evaluate representativeness of the trends derived from GB station records in the middle and upper stratosphere using SBUV data was done as a part of the LOTUS activity and was discussed in the 2019 LOTUS report. Comparisons of trends derived from satellite data sub-sampled at the station location (overpass) to those derived from the relevant zonal average provide a measure of potential sampling errors when comparing satellite and GB trends (Zerefos et al., 2018; Godin-Beekmann et al., 2022). This paper continues that work by comparing trends derived from several GB and satellite records that are matched spatially. We further investigate the impact of temporal matching on trends.

The common statistical linear regression trend model used in the 2019 LOTUS Report and the 2022 update (Godin-Beekmann et al., 2022) was optimized for analyses of the zonally averaged satellite data sets. However, analyses of the GB and satellite overpass ozone profile data may require reconsideration of additional proxies and optimization methods to improve interpretation of the processes that impact ozone changes over limited geophysical regions and reduce trend uncertainties. An assessment of model sensitivities to uncertainties in the volcanic aerosols, solar cycle, QBO, El Nino Southern Oscillation (ENSO) and other mechanisms also need to be considered in the GB and satellite overpass record trend analysis. The localized time series for the assessment of dynamical and chemical proxies can improve attribution of ozone variability, especially in the lower stratosphere, thus reducing uncertainties in the derived trends. This paper provides an assessment of uncertainties in the derived trends from the NOAA ground-based, ozonesonde and SBUV/OMPS (zonally averaged and overpass) records and reports improvements in the Multiple Linear Regression (MLR) trend uncertainties with addition of proxies representing interannual dynamical variability or long-term changes in atmospheric circulation. Ability of the ground-based and ozonesonde records in capturing semi-global ozone changes is evaluated by comparing trends derived from the satellite overpass and zonally averaged records

In the LOTUS report, the ozone trends were analyzed at low and middle latitudes, with a focus on the upper and middle stratosphere. This paper includes middle and low latitude trends assessed in the lower stratosphere and thus offers an opportunity to test the additional proxy of the tropopause pressure (Thompson et al., 2021).

2 Data

2.1 Umkehr and Ozonesonde Records at NOAA

The Dobson Ozone Spectrophotometer has been used to study total ozone since its development in the 1920s (Staehelin et al., 2018). Dobson records are regularly used in satellite record validation (Bai et al., 2015; Koukouli et al., 2016; Boynard et al., 2018) and the development of global combined ozone data records (Fioletov et al., 2008; Hassler et al., 2018). The NOAA Dobson ozone record was homogenized in 2017 to account for inconsistencies in past calibration records, data processing methods and selection of representative data (Evans et al., 2017). NOAA Dobson instruments at 4 stations and MeteoSwiss at Arosa/Davos also measure Umkehr profiles, which are derived as partial column ozone amounts in ~5 km layers. Profiles are derived using an optimum statistical inversion of Dobson measurements taken continuously at different solar zenith angles (SZAs) (Petropavlovskikh, 2005; Hassler, 2014). These Umkehr data were recently homogenized to assure the removal of small but significant instrumental artifacts that can impact the accurate detection of stratospheric ozone trends (Petropavlovskikh et al., 2022, Maillard Barras et al., 2022). This study focuses on Umkehr records from the MeteoSwiss station of Arosa/Davos, Switzerland, and on Umkehr records from the NOAA stations of Boulder, Colorado, Mauna Loa Observatory (MLO), Hawaii, Lauder, New Zealand, and the Umkehr record from Observatory de Haute Provence (OHP), France. NOAA/GML for Umkehr data means that the NOAA optimization process was applied to the operational records (N-values) prior to the retrieval of ozone profiles. The source data used in this study are available at https://gml.noaa.gov/aftp/data/ozwv/Dobson/AC4/Umkehr/Optimized/. See **Table 1** for details on the GB datasets, locations, source of data and temporal extent of data used. Umkehr measurements are typically made twice per day when there is no cloud obstruction.

The ozonesonde instrument has been flown at 4 NOAA stations since the 1980s. Evolving instrumentation and standard operating procedures led to the development of data homogenization methods by NOAA and the international community (i.e., ASOPOS-1, Smit, 2014) to resolve record inconsistencies in the NOAA (Sterling et al., 2018), Canadian (Tarasick et al., 2016) and SHADOZ (Southern Hemisphere Additional Ozonesondes) networks (Witte et al., 2017; Witte et al., 2018). The effort was extended in the ASOPOS-2 (Smit et al., 2021) activity and included a larger group of stations that are part of the NDACC (Network for Detection of Atmospheric Composition Change) and WMO GAW (World Meteorological Organization Global Atmosphere Watch program) networks. The error budget for each profile is calculated and included in the archived files (Sterling, 2018). Modern ozonesonde instruments measure ozone at high vertical resolution, on the order of 100 m (Thompson et al., 2019) depending on the balloon accent velocity and the time response of the instrument.

The sondes constitute an essential component of satellite calibration and cross-calibration (Hubert et al., 2016), verification and improvement of climate chemistry and chemistry-transport models (Wargan et al., 2018; Stauffer et al., 2019). The Dobson total ozone, Umkehr and ozonesonde profile records provide key measurements for upper and middle stratospheric ozone trend calculations, and are part of the NOAA benchmark network for stratospheric ozone profile observations (SPARC/IO3C/GAW, 2019; Godin-Beekmann et al., 2022; WMO, 2022).

The ozonesonde data are used for trend analyses from OHP, Boulder, and Lauder stations where we have Umkehr observations. Ozonesondes are launched at Hilo, Hawaii, which is nearly co-located with MLO. Ozonesonde data for the Arosa/Davos panel are selected from Hohenpeißenberg (HOH), Germany station that is in close vicinity to Arosa/Davos station. Sonde measurements are typically measured once or twice per week, varying somewhat with station operational procedures.

Data for the NOAA GML ozonesonde records are publicly available from the NOAA Global Monitoring Lab (GML) at https://gml.noaa.gov/aftp/data/ozwv/Ozonesonde/. We use the '100 Meter Average Files' in each station directory. Other sonde datasets used in this study are also available at several other data centers including the World Ozone and Ultraviolet Radiation Data Centre (WOUDC, www.woudc.org), Network for the Detection of Atmospheric Composition Change (NDACC, www.ndacc.org**)** data centers, or at the Harmonization and Evaluation of Ground-based Instruments for Free-Tropospheric Ozone Measurements (HEGIFTOM, https://hegiftom.meteo.be/) archive. Table 1 denotes the source of each dataset used in this study.

Deleted:)

Deleted:)

Deleted: Data from

Deleted: the altitude

The ozonesonde data is of significantly higher vertical resolution (even when used as 100 m averages) than the Umkehr data layers of approximately 5000 m. In order to create a dataset with comparable resolution, we use the Umkehr averaging kernels (AK) to smooth the sonde data. Details appear in Appendix A. We cap the sonde profile at Umkehr layer 5 (16–32 hPa) as there is not sufficient sonde information at higher altitudes to meet the requirements of the AKs for layers 6 and above. We further match the ozonesonde data to the dates when both Umkehr and sonde data are available using ± 24 hours to find a match, then generate the ozonesonde monthly mean. Appendix D explores the impact of temporal sampling on trends. The final matched dataset, with AK averaging, is publicly available at https://gml.noaa.gov/aftp/ozwv/Publications/2023_Umkehr_Ozone_Trends_Paper/.

Location	**WOUDC Station #**	**Instrument**	**Date Range used in trend calculations**	**Source**
Arosa/Davos Arosa, Switzerland (46.8° N, 9.7° E) Davos, Switzerland (46.8° N, 9.8° E)	035	Umkehr	1980 – 2018 2018 – 2020	Optimization by NOAA/GML
Hohenpeißenberg (HOH), Germany (47.8° N, 11.0° E)	099	Ozonesonde	1980 – 2020	NDACC
Observatory de Haute Provence (OHP), France (43.9°N, 5.8° E)	040	Umkehr	1983 – 2020	NOAA/GML
		Ozonesonde	1991 – 2020	NDACC (same as HEGIFTOM)
Boulder, Colorado (40.0° N, 105.3° W)	067	Umkehr	1980 – 2020	NOAA/GML
		Ozonesonde	1980 – 2020	NOAA/GML - 100m average data
Mauna Loa Observatory (MLO), Hawaii (19.5°N, 155.6°W)	031	Umkehr	1982 – 2020	NOAA/GML
Hilo, Hawaii (19.7° N,155.1° W)	109	Ozonesonde	1982 – 2020	NOAA/GML - 100m average data

Lauder, New Zealand (45.0°S, 169.7°E)	256	Umkehr	1987 – 2020	NOAA/GML
		Ozonesonde	1987 – 2020	NDACC

[*] Note, data from the “Corrected Ozone partial pressure” column is used for trend analyses

Table 1: GB datasets, location, instrument type, temporal extent, data record source. For the trend calculations we remove data during volcanic periods from 1982–1984 and 1991–1994.

2.2 The NOAA Cohesive (COH) Station Overpass Ozone Profile Datasets

NASA and NOAA have produced satellite measurements of ozone profiles through the Solar Backscatter Ultraviolet (SBUV) on the sequence of Polar Orbiting Environmental Satellites (POES) since 1978. This measurement series is extended with the related Ozone Mapping and Profiler Suite (OMPS) nadir profiler (NP) instruments using similar measurement techniques and retrieval algorithms. These combine to provide nearly 45 years of continuous data (1978 – present). This single instrument type dataset eliminates many homogeneity issues including varying vertical resolution, or instrumentation differences. Version 8.6 SBUV data incorporates additional calibration adjustments beyond the Version 8 release (McPeters et al., 2013). Small but evident biases remain (Kramarova et al., 2013a). Several methods have been historically used to combine these datasets into a continuous series. The NASA MOD version 1 dataset based on SBUV and OMPS v8.6 (Frith, 2014) combines data from all available satellites with no modification or bias adjustments. NASA has developed an alternate processing for the SBUV and OMPS data (v8.7) which incorporates new calibrations at the radiance level, and updated a priori with improved troposphere. Additionally, the a priori is chosen to be representative of the local solar time of the measurement. MOD v2 is based on the v8.7 processing (Frith et al., 2020), and further applies an adjustment to the v8.7 data to shift all measurements to a nominal measurement time of 1:30 PM local time.

The NOAA SBUV/2 and OMPS Cohesive dataset (further referred to as COH) combines data from the SBUV/2 and OMPS instruments using NASA’s version 8.6 for the SBUV/2 data and NOAA/NESDIS version 4r1 for the OMPS Suomi National Polar-orbiting Partnership (SNPP) data. This dataset uses correlation-based adjustments providing an overall bias adjustment plus an ozone-dependent factor (Wild et al., 2016) to moderate the remaining biases between instruments in the series. The resulting profile product is a set of daily or monthly zonal means and is publicly available at https://ftp.cpc.ncep.noaa.gov/SBUV_CDR. Zones are 5° wide in latitude, identified by the central latitude (2.5°, 7.5°, etc.). Contributing satellites and their period of use is shown in Table 2.

Satellite	**Dates**
Nimbus 07	10/1978 – 5/1989
NOAA 11 (ascending)	6/1989 – 12/1993
NOAA 09	1/1994 – 6/1997
NOAA 11 (descending)	7/1997 – 12/2000

Formatted: Left, Indent: Left: 0.25"

Formatted Table

NOAA 16	1/2001 – 12/2003
NOAA 17	1/2004 – 12/2005
NOAA 18	1/2006 – 12/2010
NOAA 19	1/2011 – 12/2013
SNPP	1/2014 – present

Table 2: Satellite mapping for COH data series.

A previous version of this dataset using OMPS v3r2 has been used in climate reviews and trend studies (Godin-Beekmann et al., 2022; Weber et al., 2022a, 2022b) including Chapter 3 of the WMO Ozone Assessment (Hassler et al., 2022). Appendix B examines the differences between the data versions. The impact on trends is limited to less than 1% per decade, well within the precision of the trend results.

We create the overpass data at a ground station by collecting all profiles from a satellite within a ±2/20° latitude/longitude box centred on the station. The box size is chosen to ensure that one to four points are found per day. Fewer points are found if the orbit passes directly over the station; more points are found if the orbits straddle the station. The collected profiles are inverse-distance weighted to the station location and averaged. COH style adjustments are applied (Wild et al., 2016) creating a COH overpass time series from 1978 to present. This dataset is available on the NOAA website at https://ftp.cpc.ncep.noaa.gov/SBUV_CDR/overpass.

Figure 1 shows the ozone anomaly time series for the 40–45° N zonal average data, and for the data at 3 stations in or near the zone. Anomalies are calculated with respect to the zonal average climatology. The series shown are for the layer data with the bottom pressure of the layer displayed on the right side of the graph. This depiction retains the information about the relative differences between the stations and the zonal average. In the mid-stratosphere (25–10 hPa) the biases between the stations are most pronounced with Arosa/Davos usually showing less ozone and Boulder usually showing more ozone. At the uppermost layers (1 and 4 hPa), and the lowest layer (41 hPa) the bias between stations is reduced. The anomalies for Arosa/Davos and OHP, which are geographically closer than Boulder, are often nearly anticorrelated with the Boulder anomalies especially in the second half of the year. Indeed at 16 hPa in particular, one can see that often the Boulder anomalies are positive when the Arosa/Davos and OHP anomalies are negative.

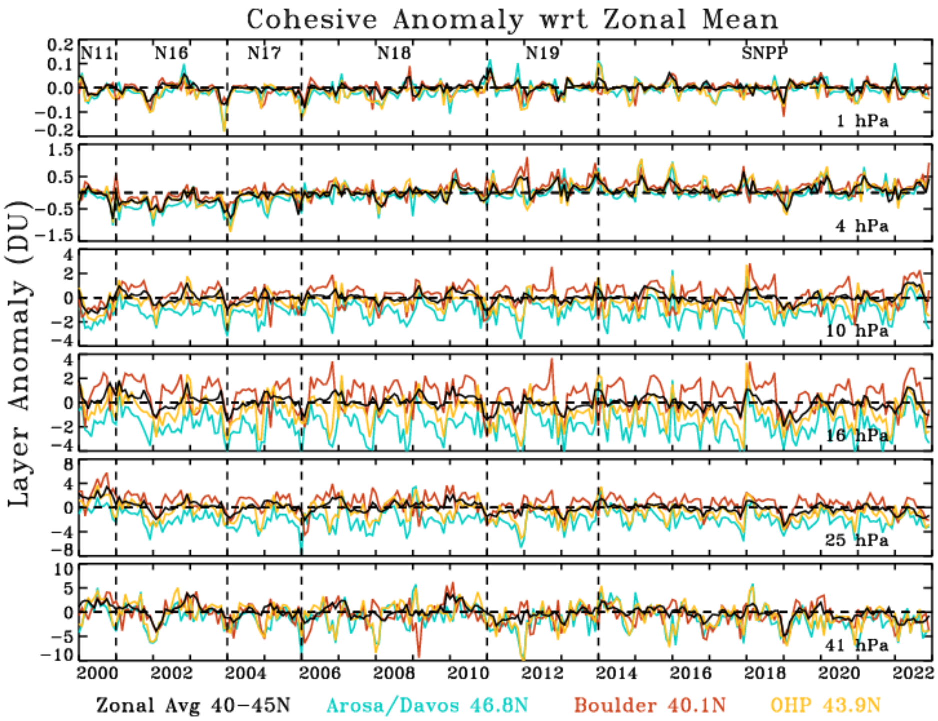


Figure 1. Monthly ozone anomaly relative to the zonal mean monthly averages. This process leaves intact the trend for each site and the zone, and accentuates the differences between the station values since all anomalies are referenced to the zonal product. Evident at 4 hPa is a positive trend from 2002 to 2013, then a levelling out after.

Figure 2 also shows the anomalies for the 40–45° N zonal average with the station anomalies, but each anomaly is now created using the climatology derived from each separate dataset. This removes the bias between the stations and the zones. At 1 hPa, Arosa/Davos appears to display the most variation (largest peaks and dips) in the anomalies. Since the anomaly for each site is now based on the seasonality of each site's data the structure in the anomalies is more uniform. For example, now at 16 hPa, the difference between Boulder and the two sites Arosa/Davos and OHP in the latter half of the year is removed. In 2012, where the Boulder anomaly was positive with respect to the zonal average seasonal value, and the Arosa/Davos and OHP sets were negative with respect to the zonal seasonal average, all are now of the same sign with respect to their own seasonal averages. Nonetheless, there are events where one station shows opposite anomalies to the other two, for example early 2009 at 41 hPa, when the Boulder anomaly is negative, and Arosa/Davos and OHP are positive. Thus, it is noted that when comparing daily or monthly data values from GB and satellite data, the overpass data will reveal a different structure than the zonal data. The trend calculations in this paper are based on the datasets of Fig. 2, where the seasonal behavior is removed using the station seasonality.

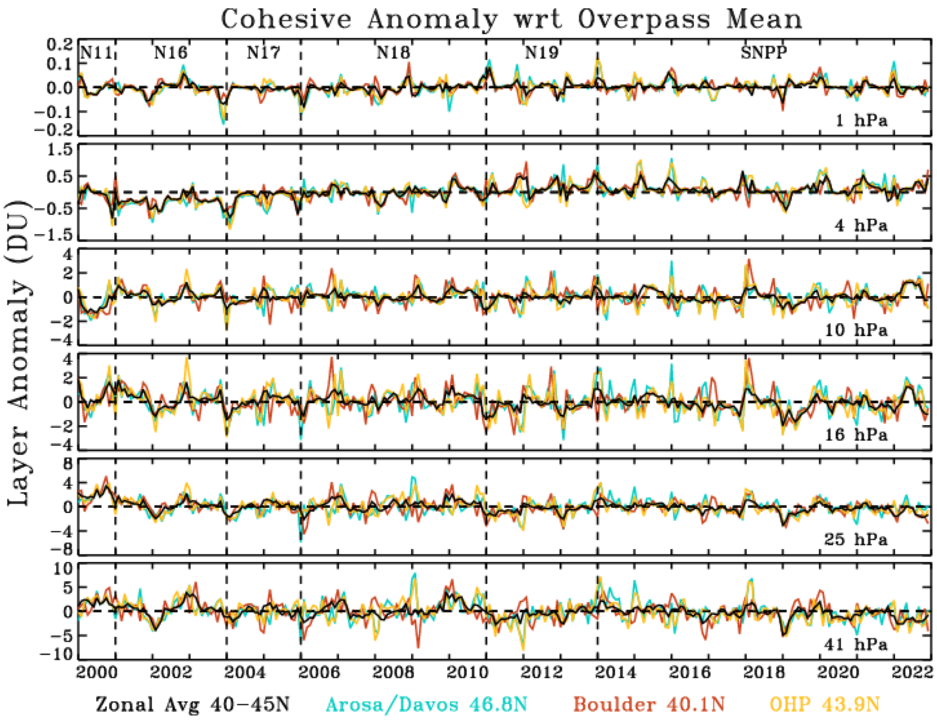


Figure 2. Monthly ozone anomaly relative to the monthly climatology for each station overpass dataset. This process leaves intact the trend for each site and the zone, and shows the consistency among the stations when each station climatology is removed. This dataset is used for the trends calculations. Evident at 4 hPa is a positive trend from 2002 to 2013, then a leveling out after. Trends are run on this dataset.

The COH overpass and zonal datasets have a similar vertical granularity as the Umkehr dataset, but use somewhat different pressures for the demarcation of the top and bottom of each layer. Since no additional smoothing is required, we simply use interpolation and integration to convert the COH layer profiles to the Umkehr layers. We exclude layers 1 to 4 since there is little sensitivity in SBUV and OMPS NP in these layers (Kramarova et al., 2013b). The overpass monthly-mean dataset in this study uses all COH data matched to dates when Umkehr data also exists. This dataset is publicly available at https://gml.noaa.gov/aftp/ozwv/Publications/2023_Umkehr_Ozone_Trends_Paper/. Appendix D explores the impact of temporal sampling on trends.

This study also uses a specialized zonal monthly-mean COH product which is the average of all daily profiles with an Umkehr match at the associated GB station. Zones used for most stations are the 5° wide zone which includes the geographic station latitude (Arosa/Davies: 47.5° N, OHP: 42.5° N, MLO: 17.5° N). Boulder and Lauder, however, are located directly on the border of two zones, so the zonal product in this study is the mathematical average of the two adjacent zones (Boulder: 37.5° N and 42.5° N, Lauder 42.5° S and 47.5° S).

3 Methods

3.1 LOTUS Model overview - the Reference Model

The Long-term Ozone Trends and Uncertainties in the Stratosphere (LOTUS) activity is a project of SPARC (Stratosphere-Troposphere Processes and their Role in Climate) and has produced a statistical Multiple Linear Regression (MLR) model called the LOTUS model (https://usask-arg.github.io/lotus-regression/index.html). The 2019 LOTUS report (SPARC/IO3C/GAW, 2019) and update (Godin-Beekmann et al., 2022) have quantified stratospheric ozone trends and evaluated their uncertainties. The LOTUS model is a general-least-squares approach MLR model. This study uses version l (v 0.8.0) with the independent linear trends (ILT) configuration. The independent linear terms represent the ozone depletion period (pre-1997), the ozone recovery period (post-2000) and an optional gap period (1997–2000). We will call the terms "pre", "post" and "gap" for short. The version 0.8.0 adds an option to enforce continuity across the gap period which is used in this study. The regression uses an interactive procedure (Cochrane and Orcutt, 1949) and the autocorrelation coefficient is adjusted with each iteration. The covariance matrix is modified accordingly to account for measurement gaps (Savin and White, 1978).

The LOTUS model (further referred as reference model in this study) is written here:

$$\hat{y}(t,z) = \beta_0(t,z)C_{pre}(t) + \beta_1(t,z)C_{post}(t) + \beta_2(t,z)Linear_{pre}(t) + \beta_3(t,z)Linear_{post}(t) + \sum_{i=4}^{n}\beta_i X_i(t,z) + \epsilon(t,z) \quad (1)$$

where $\hat{y}(t,z)$ is the estimated ozone at time ***t*** and altitude ***z***; β are the fitted coefficients of the model; the residual term, ***ε(t, z)*** is the difference between the LOTUS model and the input data. $\boldsymbol{C_{pre}}$ and $\boldsymbol{C_{post}}$ are the constant terms as defined by:

$$Constant_{pre} = \begin{cases} 1 & \text{for } t < \text{1997-Jan} \\ 1 - mt & \text{for 1997-Jan} \leq t < \text{2000-Jan} \\ 0 & \text{for } t \geq \text{2000-Jan} \end{cases}$$

$$Constant_{post} = \begin{cases} 0 & \text{for } t < \text{1997-Jan} \\ mt & \text{for 1997-Jan} \leq t < \text{2000-Jan} \\ 1 & \text{for } t \geq \text{2000-Jan} \end{cases}$$

where m=0.029135 and t = month starting in January 1980 and ending in December 2020. Indeed, the constant terms are only constant in the "pre" and "post" periods. The 3-year "gap" period is represented by a line of slope m connecting the two constant (pre and post period bias) terms.

The linear terms of the model are defined as:

$$linear_{pre} = \begin{cases} mt - b & \text{for } t < \text{1997-Jan} \\ 0 & \text{for } t \geq \text{1997-Jan} \end{cases}$$

$$linear_{post} = \begin{cases} 0 & \text{for } t < \text{2000-Jan} \\ mt & \text{for } t \geq \text{2000-Jan} \end{cases}$$

where m=0.008487, b = -1.700240, and t = month starting in January 1980 and ending in December 2020.

Natural variability is a complicating factor in deriving trends associated with the changes in the ozone depleting chemistry. LOTUS fits predictor variables as proxies for natural variability to the ozone data so that one can interpret the resulting linear trend as a trend due to the changes in chemistry. The summation term is the summation of the predictors used as a proxy for the dynamical induced ozone variability.

The natural variability proxies in the LOTUS model v 0.8.0 are Aerosol Optical Depth (AOD), El Nino/ Southern Oscillation (ENSO), and the Quasi-Biennial Oscillation (QBO) in the form of the first two principal components (also known as an empirical orthogonal function analysis). The data sources for each are described in Table 3.

Large sulfur dioxide (SO_2) levels reaching the lower stratosphere following major volcanic eruptions (i.e. El Chichon, Pinatubo or Hunga) can impact the validity of ozonesonde measured values (Yoon et al., 2022). However, SO_2 is not long-lived gas and is soon converted to sulphate aerosols that can alter observations by ozone remote sensing systems. Both Umkehr and satellite ozone profiles from SBUV and OMPS are highly uncertain and/or biased because of high aerosol load during volcanic eruptions (DeLuisi et al, 1989; Petropavlovskikh et al., 2005, 2022; Bhartia et al, 1993, Torres et al., 1995, Bhartia et al, 2013). It is recommended that the data for 2 to 3 years after the El-Chichon and Pinatubo large volcanic eruptions should not be used in trend analyses. Therefore, we exclude data during the volcanic periods (1982–1983 and 1991–1993) from the analyzed time series. Moreover, this study is focused on the linear trend analyses after 2000 when there are no large stratospheric aerosol perturbations that significantly influence stratospheric ozone variability over the middle latitudes and therefore impact trend and uncertainty estimates. Since we have eliminated the data during the volcanic period, this study does not include the AOD proxy in the calculations.

We define the 'reference' model (RM) as the proxies most commonly used for the dynamical proxies which is equivalent to the LOTUS model v 0.8.0 minus the AOD term. The representative equation is:

$$\sum_{i=4}^{n} \beta_i(t,z)X_i(t) = \beta_4(t,z)QBO_A(t) + \beta_5(t,z)QBO_B(t) + \beta_6(t,z)ENSO(t) + \beta_7(t,z)Solar(t) \quad (2)$$

The Quasi-biennial Oscillation (QBO) is derived from the Singapore radiosonde profiles (1979–2020) that detect variability in the direction of the tropical winds in the lower stratosphere. It also shows that zonal wind variation propagates downward with an average period of ~28 months [Wallace, 1973]. The principal component analysis of the 100–10 hPa zonal winds can describe the majority of the wind variability. The reference model (and LOTUS v 0.8.0) use the two leading modes of the calculated empirical orthogonal functions (EOF) for trend analyses [Wallace et al., 1993].

The El Niño/ Southern Oscillation (ENSO) is a periodic mode of climate variability of the atmosphere and sea surface temperatures associated with the equatorial Pacific Ocean with periods ranging from 2–8 years. The Multivariate ENSO index (MEI) is produced by the NOAA Physical Sciences Laboratory and is derived from the EOF analysis of sea surface temperature, sea level pressure, outgoing terrestrial radiation, and surface winds in the area of the Pacific basin from 30° S to 30° N and from 100° E to 70° W (Wolter and Timlin, 2011). Temperature anomalies in the troposphere with corresponding stratospheric temperature anomalies during El Niño/ La Niña events modulate the tropical upwelling of the Brewer-Dobson circulation (BDC) and thus the meridional transport of ozone in the stratosphere. (Diallo et al., 2018).

The solar cycle is the 11-year periodic cycle of solar activity and solar irradiance that reaches the Earth's atmosphere. The change in UV radiation that is absorbed by the atmosphere, most notably in the upper stratosphere, leads to changes in atmospheric temperature and the photochemistry which produces ozone. (Lee and Smith, 2003). The 10.7 cm solar radio flux data is used as the proxy for the solar cycle in the LOTUS model.

Seasonal components in the form of Fourier harmonics were added into the LOTUS model with version 0.8.0. Godin-Beekman et al. (2022) showed in their Fig. 7 that the model fit for the ozone profile satellite and model records is

Deleted:

Deleted: after

Deleted: ozone

Deleted: retrievals

Formatted: Subscript

improved by adding seasonal components to the proxies, increasing the adjusted R-squared (R^2) from 0.3 or less to 0.3 to 0.5. The seasonality and relevant contributions of some predictor's variables are compensated in this study by adding the seasonal components to the fitted predictors. Seasonal components are represented in the model by sine and cosine functions with periods of 12 and 6 months that describe the variability of the proxies on these timescales. So, for each fitted predictor in the model

$$\beta_i X_i(t,z) \text{ where } i > 1$$

a seasonal variation in the form of Fourier components is added as follows:

$$\beta_m(t,z) = \beta_{m,0}(z) + \sum_{i=1}^{2} \beta_{m,1,i}(z) sin(\frac{2\pi it}{12}) + \sum_{i=1}^{2} \beta_{m,2,i}(z) cos(\frac{2\pi it}{12})$$

3.2 The Extended Model - Adding Predictors

Recent publications (i.e. Petropavlovskikh et al., 2019; Szelag et al., 2020; Godin-Beekmann et al, 2022; Millan et al. 2024) highlight the need to reduce the trend uncertainties in the lower stratosphere (LS). There is still a discrepancy between modeled and observed ozone trends in the LS but large uncertainties make comparisons difficult. In this study, we test additional predictors in the model to account for dynamical variability of ozone in the stratosphere, thus improving the model performance and reducing the uncertainty of the trends. The argument for additional predictors is that the LOTUS model was developed for the regression of zonally averaged ozone data, which reduces some variability that might be impacting the ground-based records on regional bases. Impact of additional proxies in trend analyses were reported in other publications (Weber et al, 2022a, Bernet, 2023 and references therein), and were mostly found to improve the statistical model fit at high latitudes where the impact of the descending branch of the Brewer-Dobson circulation and Arctic/Antarctic oscillations has contributed to additional variability in stratospheric ozone records.

In what we define as the 'extended' model, we add single additional predictors (one at a time) in the model as such:

$$\sum_{i=4}^{n} \beta_i(t,z) X_i(t,z) = \beta_4(t,z) QBO_A(t) + \beta_5(t,z) QBO_B(t) + \beta_6(t,z) ENSO(t) + \beta_7(t,z) Solar(t) + \beta_8(t,z) X_{predictor}(t,z)$$

The fitted predictors contain Fourier components, like in the reference model, to allow for seasonal variation.

We test the following additional predictors as described below to assess the impact on trends and uncertainties:

- Quasi-Biennial Oscillation (QBO): Two notable disruptions to the otherwise relatively periodic QBO have occurred during the study period: 2015–2016 and in 2020 (Diallo, et. al 2022). Two additional leading modes of the calculated empirical orthogonal functions (EOF) are tested to improve the trend model fit during the anomalous QBO years.
- Arctic/Antarctic Oscillation (AO/AAO): the pattern of surface air pressure anomalies in the polar region and certain mid-latitude regions. The AO/AAO has strong correlations (Lawrence et al 2020) with stratospheric ozone through the strength of the polar vortex. The positive phase of the AO or AAO in the winter months is associated with low activity in the vertically propagating planetary Rossby waves, a strong polar vortex, a

Formatted: Superscript

low vortex wavenumber, and low stratospheric temperatures. Thus, the positive (negative) phase of the AO/AAO is correlated to low (high) ozone anomalies especially in the winter months (Lawrence et al, 2020).

- North Atlantic Oscillation (NAO): Similar to the Arctic Oscillation, this is a pattern of surface air pressure anomalies between certain regions in the high altitudes of the North Atlantic Ocean. This index is calculated by the pressure difference between the Azores high and the subpolar low.
- Eddy Heat Flux (EHF): The flux of heat through a zonal plane by transport due to the Brewer-Dobson circulation, here averaged from 45–75° N/S (use EHF S for Lauder only). This represents the planetary wave activity that drives transport of ozone.
- Tropopause Pressure (TP): Pressure level of boundary between the troposphere and the stratosphere. In this study, we use the monthly mean pressure level of the tropopause from the NOAA National Centers for Environmental Prediction (NCEP) reanalysis product. As the troposphere warms due to release of GHGs and the stratosphere cools due to ODSs destroying stratospheric ozone, the tropopause is rising (Meng et al., 2021). Thompson et al, (2021) and Stauffer et al., (2023) found that the lower stratospheric ozone trends in tropics become slightly positive when recomputed with respect to the tropopause height (which has its own trend). This finding indicates that ozone depletion in the lower stratosphere (i.e. Ball et al., 2020) is driven by climate-change-related changes in transport and mixing in the lower stratosphere. Therefore, we are testing the TP proxy in the model to account for non-chemical ozone losses in order to assess chemical attribution of ozone trends.
- Equivalent Latitude (EqLat): Geographical latitude of the isoline encircling the area of equal Potential Vorticity (PV) (Lary et al, 1995). The EqLat normalizes the range of PV values that change with season and interannually and makes it convenient for interpretation of ozone variability and trends (i.e. Wohltmann et al 2005). The dataset was generated from GMI CTM analyses (private communications with Susan Strahan, June 2021) for each ground-based station overpass criteria (latitude and longitude envelope, see above) and at several altitude levels coincident with Umkehr ozone profile layers. Appendix C discusses a COH dataset based on EqLat instead of geometric latitude. No advantage was found by using the EqLat coordinate system for the COH zonal dataset.

Source datasets for all predictors in the reference and extended models are shown in Table 3.

Predictor	Description	Source
ENSO	**El Nino/Southern Oscillation**	**Monthly Mean Multivariate ENSO Index https://psl.noaa.gov/enso/mei.old/**[1]
Solar	**Solar 10.7cm flux**	**https://spaceweather.gc.ca/forecast-prevision/solar-solaire/solarflux/sx-5-en.php**

QBO	**Quasi-Biennial Oscillation**	**Principal Component Analysis of the Monthly Mean Zonal Wind https://www.geo.fu-berlin.de/met/ag/strat/produkte/qbo/qbo.dat**
AOD	**AOD is included in the LOTUS model, but not used in this study**	
AO	Arctic Oscillation, Monthly Mean index	http://www.cpc.ncep.noaa.gov/products/precip/CWlink/daily_ao_index/monthly.ao.index.b50.current.ascii
AAO	Antarctic Oscillation, Monthly Mean index	https://www.cpc.ncep.noaa.gov/products/precip/CWlink/daily_ao_index/aao/aao.shtml
NAO	The North Atlantic Oscillation, monthly mean index	https://www.cpc.ncep.noaa.gov/products/precip/CWlink/pna/norm.nao.monthly.b5001.current.ascii.table
EHF	Eddy Heat Flux	Cumulative Mean (from September to April) of Heat Flux at 100 hPa from MERRA2 reanalysis averaged over 45–75° N (45–75° S for Lauder), deseasonalized. It is kept constant from April to Sep. https://acd-ext.gsfc.nasa.gov/Data_services/met/ann_data.html
TP	Tropopause Pressure	Monthly Mean NCEP-NCAR reanalysis (Kalnay et al., 1996); Tropopause pressure at the lat/lon of each station, deseasonalized. ftp://ftp.cdc.noaa.gov/Datasets/ncep.reanalysis.derived/tropopause/pres.tropp.mon.mean.nc
EqLat	Equivalent Latitude	Monthly Mean equivalent latitude derived from MERRA2 -GMI CTM potential vorticity (PV) contours on 31 potential temperature surfaces

		[Susan Strahan, private communication, 8/24/2022]. The PV at each station is determined by a 1/distance weighted average of the values in a ± 2 lat, ± 2 lon grid, then converted to EqLat on the Umkehr layers.

Table 3: List of predictors either previously used (bolded) in the LOTUS 0.8.0 (reference) model and additional predictors evaluated in this study for a future use in the extended LOTUS trend regression model. Note, two components of the QBO predictors were used in the reference model (i.e. Godin-Beekmann et al., 2022). We added two more components in the extended model for tests described in this paper.

[1] Since the incorporation of the ENSO index into the LOTUS model, NOAA GSL has updated the index to v1.2. https://psl.noaa.gov/enso/mei/. However, for consistency with results from the Godin-Beekmann (2022) paper we use the old MEI index that is part of the LOTUS v 0.8.0 package.

All proxies are used as is. No de-trending (removal of the long-term trend in proxy) is applied to the proxies. Therefore, we interpret any changes to the trends derived with additional proxies as approximations of trends driven by chemistry and transport related to climate change. These are rough approximations as some feedbacks are known to impact chemistry (e.g. changes in stratospheric temperature).

3.3 The Full Model - Combining Additional Predictors

After we have determined the impact of the additional predictors singly, we discern which predictors should be combined to constitute the 'Full Model'. Prior to selecting additional predictors for the 'Full Model', we perform correlation tests to identify any cross correlations between predictors. We select predictors that are not highly correlated (less than +/- 0.2) to ensure that all predictors are largely independent. We use the square of the Pearson correlation coefficient R^2 for each pair of the predictors to test our assumptions. We find that ENSO, Solar, QBO (1,2,3,4), AAO, AO, EHF (N and S), and TP (at each station) have correlations less than +/- 0.2 (with the exception of R^2 = 0.3 for EHF (N) and AO). Therefore, any of these predictors can be combined in the 'Full Model'. We find that NAO has a correlation of .38 with AO so we do not use these two predictors in the same model.

We also test the independence of EqLat proxies calculated at several geographic locations (defined by the latitude and longitude of each Umkehr station) and by selecting a proxy at several altitude levels centered in the middle of Umkehr layers 3–9. We find that the R^2 between the TP and EqLat in the lower stratosphere (Umkehr layer 3) can be large but anticorrelated -0.7 (Boulder), moderate 0.4 (MLO and Lauder), while close to zero at Arosa/Davos and low at OHP (-0.2). In the middle and upper stratosphere, the R^2 varies from -0.5 to -0.4 (MLO), 0.2 to 0.3 (Arosa/Davos and OHP), 0.5 to 0.6 (Boulder), and 0.4 to 0.7 (Lauder). EqLat has mostly low correlations (< ± 0.3) with all other proxies except for higher correlations with QBO B in layers 5 (-0.3) and 6 (-0.4), and QBO A in layer 7 (0.3) at MLO; and with AO in layer 8 (0.3) at OHP and Arosa/Davos. Also, EqLat has no correlation with the TP proxy in layer 4 in Boulder, in layer 9 at Lauder, and in layers 8 and 9 at OHP. Since there are occasional high correlations between EqLat and TP proxies, we do not use them together in the 'Full trend Model'.

Formatted: Superscript

Formatted: Superscript

Formatted: Superscript

Formatted: Superscript

4 Results

4.1 Reference Model Trend Results

First, we discuss the reference model trends derived from the COH overpass, Umkehr and ozonesonde records at 5 geographic locations. All datasets are deseasonalized with a climatology computed from a subset of data taken from 1998–2008 prior to the trend analysis. Trend results are presented in Fig. 3 and organized in 5 panels. Each panel shows trends at selected pressure/altitude levels detected from Umkehr (green), COH (orange) and ozonesonde (blue) records at Arosa/Davos, OHP, Boulder, MLO/Hilo and Lauder ground-based stations. Ozonesonde data for the Arosa/Davos panel are selected from Hohenpeißenberg, Germany station that is in close vicinity to Arosa/Davos station. We show trends for layers where the measurement is of highest quality: Umkehr (layer 3 through 8), COH (layers 5 through 9) and ozonesonde (layers 3 through 5) records.

The Umkehr data used in this analysis is the monthly mean of all available Umkehr data (one or two measurements per day). The sonde and COH monthly means use only those profiles that have corresponding Umkehr measurements on that date. We explore the impact of temporal sampling on trends in Appendix D. For COH with the Umkehr matched data (see Figure A12), trends are slightly larger at OHP but well within the error bars. At all other stations the COH trends are not impacted by sampling. At OHP the ozonesonde trends matched to Umkehr (see Figure A13) are slightly larger at layer 4 only and well within the error bars; while at Lauder in layers 4 and 5 trends are smaller, but barely within the error bars.

Figure 3 shows that in the upper (above 10 hPa) stratosphere, Umkehr (black) and COH (orange) trends are positive and agree within the error bars (+/- 2 standard errors). The exception is found at 8–2 hPa pressure level over the Lauder station, where Umkehr trends are near zero and COH trends are ~ +3–4 %/decade. The error bars show +/- 2 standard errors, and the fact that they do not overlap suggests that the differences in trends are statistically significant. This could be related to the relatively large uncertainties in the instrumental corrections applied to homogenize the Umkehr record (Petropavlovskikh et al, 2022). Björklund (2023) discusses relative drifts in Umkehr, ozonesonde, FTIR and MW ozone records over Lauder. The authors are not able to identify instrumental artifacts that may have caused the discrepancies in the co-located records, but point out that it is not related to the sampling biases.

In the middle stratosphere (60–10 hPa) agreement between Umkehr and COH is within uncertainty of the trend except at Arosa/Davos where COH trends are statistically different from Umkehr trends at 16–8 hPa. COH trends at 32–16 hPa are mostly negative (-2–3 %/decade) with the exception of Lauder where trends are near zero and similar to Umkehr trends. Umkehr trends between 32–16 hPa are close to zero. The ozonesonde trends (blue) agree with COH (orange) and Umkehr (black) trends in layer 32–16 hPa at Arosa/Davos, Boulder and MLO. However, at OHP (Lauder) the ozonesonde trends are found to be positive at +2±2.2 %/decade (negative at -3±1.5 %/decade) and significantly different from the near-zero trends seen in the COH and Umkehr results.

In the lower stratosphere (125–63 hPa), Umkehr trends vary between small positive (+1–2 %/decade at Hilo and Lauder) and negative (-2–3 %/decade at Arosa/Davos, OHP and Boulder); however, trend uncertainties are the largest (2 standard errors are 2–3 %/decade, see Table 4 below) in comparison to the middle and upper stratospheric trends. Ozonesonde trends at OHP station are positive (+2 %/decade), and negative over Lauder (-2 %/decade). They also feature large uncertainties (±4.2 %/decade at OHP) that are larger than the uncertainties found in Umkehr trends which

Deleted: I
Deleted: green
Deleted: green
Deleted: 63
Deleted: 3
Deleted: 3
Deleted: (
Deleted: 4
Deleted: 5

could be caused by the limited sampling (see Appendix D, Figure A11). Sonde trends at Hilo show negative trend values with large uncertainties. But the data in this study at Hilo is not corrected for the ozonesonde drop off after 2014 known to occur at this station (Stauffer, 2022), so the deviation from the Umkehr results at these levels may be misleading.

Figure 3 also shows trends derived from the zonal-mean COH data associated with each station (orange dashed line). These are shown for comparison with the overpass COH data (solid line) to study the impact of the spatial sampling biases on the trends. Though Figs. 1 and 2 show clear interannual differences between the records from the individual stations, and the associated zonal average, we find very small differences in trends (mostly in the upper stratosphere at middle latitude stations). Therefore, the station overpass sampling provides trends that are representative of the zonal averaged trends (Zerefos, 2018) and the discrepancies in trends between GB and satellite records do not strongly depend on the spatial sampling differences.

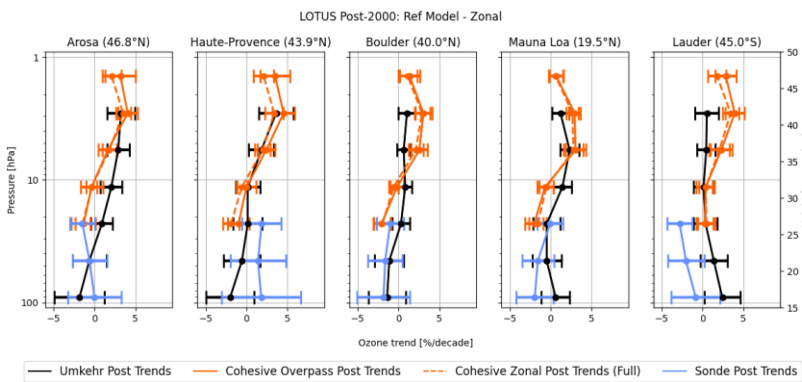
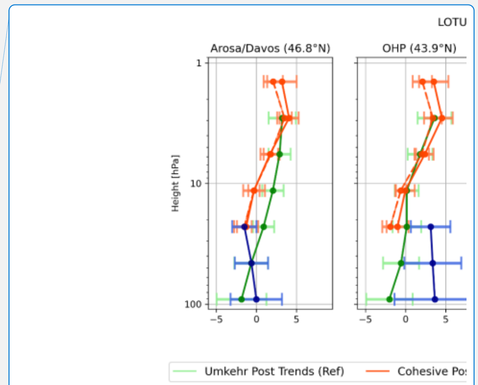


Figure 3: The 2000–2020 ozone trends are shown at 7 altitude/pressure levels. The LOTUS model v 0.8.0 is used for trend analyses. Umkehr trends (black), COH (orange) and ozonesonde (blue) are shown for 5 ground-based stations: Arosa/Davos, OHP, Boulder, MLO and Lauder (panels left to right). Ozonesonde data for the Arosa/Davos panel are selected from Hohenpeißenberg, Germany that is in close vicinity to Arosa/Davos. Trends from the zonal-mean COH data (orange dashed line) are shown for comparison with the overpass COH data (solid line). The error bars indicate ± 2 standard errors.

4.2 Standard Error of Reference Model

LOTUS Model Proxy Tests: Standard Error for Reference Model																
Height	**Umkehr**	**Arosa/Davos**			**OHP**			**Boulder**			**MLO**			**Lauder**		
(hPa)	**Layer**	**UMK**	**COH**	**SND**	**UMK**	**COH**	**SND**	**UMK**	**COH**	**SND**	**UMK**	**COH**	**SND**	**UMK**	**COH**	**SND**
1-2	**9**		0.92			0.91			0.62			0.43			0.63	
2-4	**8**	0.85	0.59		1.06	0.68		0.51	0.52		0.52	0.37		0.72	0.57	
4-8	**7**	0.69	0.59		0.77	0.54		0.41	0.52		0.58	0.62		0.57	0.66	
8-16	**6**	0.66	0.68		0.75	0.59		0.42	0.43		0.55	0.49		0.61	0.56	
16-32	**5**	0.66	0.75	0.76	0.89	0.68	1.10	0.54	0.51	0.77	0.82	0.55	0.75	0.73	0.54	0.73

Deleted: ¶



Deleted:

Formatted: English (United States)

Deleted: green

Deleted: 1.26

32-63	**4**	1.05		1.04	1.13		1.55	0.90		1.04	0.90		0.94	0.83		1.16
63-127	**3**	1.55		1.60	0.15		2.10	1.15		1.63	0.87		1.07	1.11		1.50

Deleted: 1.95
Deleted: 2
Deleted: 75

Table 4: Standard Error (SE) for the Reference model 2000–2020 trend for five ground-based station locations (Arosa/Davos, OHP, Boulder, MLO and Lauder). Results are provided for trend analyses of the Cohesive satellite (COH), Dobson Umkehr (UMK) and ozonesonde (SND) records and for Umkehr. The layers are selected to represent the best quality of data. Values of SE shown are the actual errors in DU/decade.

We will use the standard error of the linear (trend) term in Equation 1 to evaluate the success of the additional proxies to improve understanding of trend values. The standard error is an output of the regression code, and indicates the uncertainty in the trend value. Smaller Standard Errors indicate increased confidence in the trend result. We use the standard error as a metric instead of standard deviation to reduce dependence on the number of points in the trends model. The **Table 4** provides the Standard Errors for the Reference Model fit and represents uncertainty of the trend in DU of the mean ozone in each layer at the station. The standard errors of the trend detected in three co-located ozone records at each station (or in the nearby location as in case of Arosa/Davos or MLO comparisons) do not significantly differ, although in general ozonesonde errors are slightly larger than Umkehr errors most likely due to the larger sampling errors in ozonesonde monthly mean record. Also, the errors in trends detected in COH layers 5–8 are on average smaller than for Umkehr trends (with the exception of layer 7 at Boulder, MLO and Lauder) which could be explained by an overpass method that averages several satellite profiles from adjacent orbits and therefore reduces meteorological scale variability in averaged ozone data.

Deleted: fit to the data
Deleted: improvements in the model fit after additional proxies are included…

4.3 Adjusted R^2

Formatted: Superscript

The adjusted R^2 values of the 2000–2020 trends are shown in Fig. 4 and Table 5 for the data fit using the Reference model. The adjusted R^2 is a modified version of R^2 that adjusts for the number of predictors in a regression model and represents the ‘goodness’ of the model fit to the data. For COH adjusted R^2 is shown for both the overpass and the zonal datasets.

Formatted: Superscript
Deleted: R2
Deleted: R2
Deleted: R2

Though values are significantly less than the high values usually seen when comparing data that includes the prevalent seasonal variation, the adjusted R^2 values for the COH zonal mean record are similar in magnitude and vertical shape to the results of the (60°S–60°N) broadband trend analyses published in Godin-Beekmann (2022), Fig. 7 varying between 0.1 and 0.5. We designate the average values (0.3) as a threshold for satisfactory fit indicating conformance with prior LOTUS results. We indicate in bold in Table 5 adjusted R^2 values of 0.3 or greater to note achievement of that threshold and include vertical dashed line in Fig. 4 for reference.

Deleted: R2
Deleted: R2

The adjusted R^2 for the Reference model fit is slightly better for the zonal mean COH data than for the COH overpass over the Northern middle latitude stations. This is expected as much of the variability of the time series is reduced in the zonal average as compared to the station overpass data as shown in Fig. 2, and more easily explained by the typically used predictors. Indeed, the goal of this study is to determine if the additional predictors help to explain the additional variation as measured at point locations.

Deleted: R2

The model fit to the GB data is similar to the COH overpass results in the middle stratosphere (layers 5 and 6), but the model explains less ozone variability in the Umkehr records in the upper stratosphere (layers 7 and 8). In the lower stratosphere (layers 3, 4 and 5), the model fit to the ozonesonde and Umkehr records is similar with the exception of Lauder (Umkehr has larger adjusted R^2 in layers 4 and 5). The adjusted R^2 for COH overpass in layer 5 is similar to Umkehr and sonde with a larger difference at OHP. The adjusted R^2 in the lower stratosphere is less than in the middle stratosphere, which points to other processes (e.g., transport) that drive ozone variability. In this paper we investigate improvement to the trend model fit by introducing additional proxies that can improve representation of the dynamically-driven ozone variability in the stratosphere.

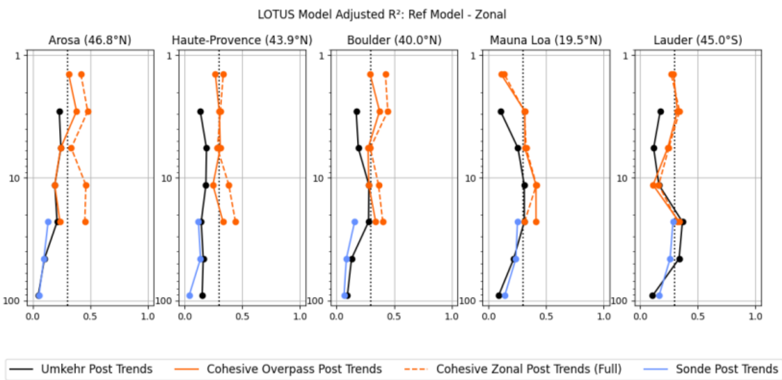


Figure 4: The adjusted R^2 is plotted as a function of altitude/pressure for the LOTUS model fit to the Umkehr (black), ozonesonde (blue), COH overpass (orange, solid), and COH zonal-mean (orange, dashed). Results are shown in 5 panels that represent trend analyses of ozone records over Arosa/Davos (Hohenpeißenberg for sondes), OHP, Boulder, MLO (Hilo for sondes) and Lauder ground-based stations.

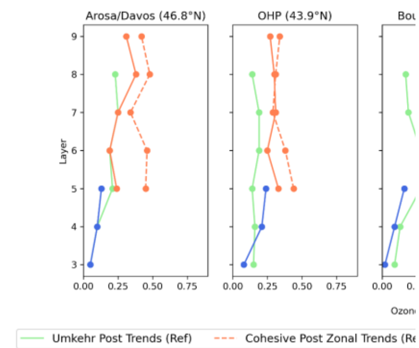
LOTUS Model Proxy Tests: Adjusted R2 for Reference Model																
Height	**Umkehr**	**Arosa/Davos**			**OHP**			**Boulder**			**MLO**			**Lauder**		
(hPa)	**Layer**	**UMK**	**COH**	**SND**	**UMK**	**COH**	**SND**	**UMK**	**COH**	**SND**	**UMK**	**COH**	**SND**	**UMK**	**COH**	**SND**
1-2	**9**		**0.31**			0.27			0.29			0.11			0.29	
2-4	**8**	0.23	**0.38**		0.14	**0.30**		0.17	**0.37**		0.11	**0.32**		0.17	**0.32**	
4-8	**7**	0.25	0.25		0.19	**0.31**		0.19	0.27		0.26	**0.32**		0.12	0.24	
8-16	**6**	0.19	0.19		0.19	0.25		0.28	0.28		**0.31**	**0.41**		0.16	0.11	
16-32	**5**	0.21	0.24	0.13	0.14	**0.33**	0.14	0.28	**0.34**	0.16	**0.31**	**0.41**	0.25	**0.37**	**0.34**	0.26
32-63	**4**	0.10		0.10	0.16		0.24	0.13		0.09	0.22		0.24	**0.34**		0.20
63-127	**3**	0.05		0.05	0.15		0.25	0.09		0.02	0.09		0.14	0.11		0.10

Table 5: Similar to Table 4, but for the adjusted R^2. Values of 0.30 and above are indicated in Bold as a threshold to indicate a satisfactory fit.

4.4 Reference Model P-Values:

The p-values are often used to evaluate statistical significance of predicted results and results labelled “significant” if they remain below a threshold of 0.05. However, Chang et al. (2021) argued as Wasserstein et al. (2019) does that all trends should be reported with their associated p-values and a thorough discussion of the certainty of trend detection

Deleted: R2
Deleted: R2
Deleted: R2
Deleted: ¶



Formatted: English (United States)
Formatted: Superscript
Deleted: green
Deleted: 24
Deleted: 1
Deleted: 08
Formatted: Superscript

as described by the p-values. Therefore, the p-values can be used for understanding the certainty of the trend. Under the IGAC TOAR activity, p-values are scored to define a consistent scale for comparison of the trends between different analyses (see Table 3, Chang et al., 2023).

LOTUS Model Reference Model P Values																
Pressure	Umkehr	Arosa/Davos			Haute-Provance			Boulder			Mauna Loa			Lauder		
(hPa)	Layer	UMK	COH	SND	UMK	COH	SND	UMK	COH	SND	UMK	COH	SND	UMK	COH	SND
1-2	9		**0.00**			**0.00**			**0.03**			0.10			**0.00**	
2-4	8	**0.00**	**0.00**		**0.00**	**0.00**		**0.05**	**0.00**		**0.02**	**0.00**		0.47	**0.00**	
4-8	7	**0.00**	**0.01**		**0.02**	**0.00**		0.12	**0.00**		**0.00**	**0.00**		0.43	**0.00**	
8-16	6	**0.00**	0.62		0.84	0.98		0.05	0.66		**0.01**	0.17		0.85	0.50	
16-32	5	0.17	0.08	**0.05**	0.87	0.15	0.12	0.58	**0.00**	0.21	0.56	**0.00**	0.93	0.61	0.62	**0.00**
32-63	4	0.55		0.57	0.62		0.41	0.21		0.12	0.61		0.11	0.10		0.07
63-127	3	0.23		1.00	0.17		0.46	0.22		0.26	0.47		0.08	**0.03**		0.58

Table 6: Similar to Table 4 but for p-values. Values of less than 0.05 (high certainty of trend detection) are shown in blue and with bolded numbers. Values between .05 and 0.1 (blue, not bolded) indicate medium certainty, between 0.1 and 0.33 (orange) - low certainty of trend detection, and above 0.33 (red) - very low certainty or no evidence of trend detection.

Table 6 provides p-values for the Reference Model. These are further used as a baseline for comparison to model fits with additional predictors. P-values of the reference model fit suggest a high certainty (p<0.05) in detected trends in the COH data in layers 7, 8 and 9 at almost all stations with the exception of the higher p-value (0.1, medium certainty) found at MLO in layer 9. Also, high certainty in derived trends is reached for COH records in layer 5 at Boulder and MLO.

Umkehr trend analyses also show high confidence in trend detection at Arosa/Davos and MLO stations in layers 6, 7 and 8, at OHP in layers 76 and 8, and in Boulder in layers 6 and 8. For the ozonesonde data the high confidence (i.e. low uncertainty) is found for Hohenpeißenberg, and Boulder trends detected in layer 5, and at Lauder in layers 4 and 5.

The medium level of the certainty (0.05<p≤0.10) is found in trends detected in layer 5 of COH ozone time series at Arosa/Davos, layer 3 of ozonesonde at MLO, and in layer 4 of ozonesonde and layer 3 of Umkehr at Lauder. Low certainty in detected trends at p-value of 0.10 (not inclusive) to 0.33 is found in Umkehr layer 3 and 5 at Arosa/Davos; in COH layer 5 and Umkehr layer 3 at OHP; in Umkehr layers 3 4 and ozonesonde layers at Boulder; in ozonesonde layer 4 and COH layer 6 record at MLO.

Highest (lowest certainty) p-values (>0.33) were found in layer 6 of COH overpass records at most stations (except for MLO where p-values are medium high). We note that the COH trends are close to zero and the uncertainty envelope crosses the zero line. Therefore, the statistical trend model cannot separate trends from zero due to unexplained high ozone variability in this layer. Similarly, near-zero Umkehr trends with relatively large SE in layer 6 at OHP and Lauder, layer 5 at all (except Arosa/Davos) stations, and in layers 3 and 4 at MLO show the same level of high p-values thus suggesting that additional proxies should be added in the trend model to assess the impacts of the natural variability and instrumental noise on trend uncertainty.

Deleted: LOTUS Model Proxy Tests: Reference Model P Va ...

Formatted: English (United States)

Formatted: Normal

Deleted: green

Deleted: yellow

Deleted: OHP,

Deleted: layer 4 of ozonesonde records at OHP,

Deleted: layer 9 COH and

Deleted: ,

Deleted: and ozonesonde

Deleted: ,

Deleted: and 7

Deleted: 3 and 4

Deleted: and

Deleted: Therefore, these results point to the trend model's inability to detect non-zero trends and account for all ozone variability in this layer

It is also important to note that the reference trend model fit to ozone in Umkehr layers 7 and 8 at Lauder has high p-values, which is related to the near-zero trends that shows large disagreement with COH trend. This difference could be caused by remaining instrumental step changes that were not fully removed during the record homogenization (Petropavlovskikh et al., 2022).

While near-zero trends and high p-values are found in the fit of the Hilo ozonesonde record in layer 5, the p-values in layer 4 show only medium p-values for near zero trends. It is possible that infrequent launches of ozonesonde observations at Hilo could create the temporal sampling bias and appear noisy. The ozonesonde record at Hohenpeißenberg has sufficiently frequent sampling (3 times per week) for successful trend analyses (Chang et al., 2020; Chang, 2023 preprint), but the p-values remain high in layers 3–4. The p-values for Umkehr fit at Arosa/Davos are in the medium to high range for layers 3, 4, 5, but somewhat smaller which could be due to non-zero trends in layers 3 and 5. The p-value difference could be also related to the different location of the ozonesonde (HOH) and Umkehr (Arosa/Davos) observations, thus the records could contain different atmospheric variability that might impact the model fit.
We will discuss changes to the p-values in the next section after we add more proxies to the trend model in an attempt to improve confidence in trend detection.

5 Trends with the Extended Model - testing the addition of single predictors

The LOTUS styled Reference Model is developed and optimized for zonal average datasets. Modeling and trend analysis for GB and satellite overpass data may improve by the addition of other proxies not used in the reference model to improve capturing processes that impact ozone changes over limited geographical regions. The Extended Model tests the addition of single predictors to see if fit statistics can be improved for GB and overpass datasets. We judge success of the Extended Model by examination of the reduction in the Standard Error of the trend term, and by evaluation of the impact on the adjusted R^2 of the model fit. Table 7 displays the change in the Standard Error of the post 2000 trend for each proxy tested determined as SE_{ref} - SE_{ext} as a percent of SE_{ref}. As such positive values correspond to the desired reduction of SE, and are highlighted in the table in blue. Low impact changes in the SE are highlighted in white, and increases in SE (negative values) are highlighted in red. It may seem unusual for the addition of proxies to increase the SE (negative values in the table) which indicates less confidence in the fit. But these SE are the uncertainty in the trend term, not in the overall model fit. The new proxies considered each have a possible trend and associated error budget for that trend. Whether the additional proxy increases trend uncertainty can depend on how well the trend of the new proxy can be characterized. The adjusted R^2 is a better indicator of the overall model improvement. Table 8 displays the adjusted R^2 for the Extended Model for each proxy tested. Values of 0.30 and above are indicated in bold as a threshold to indicate a satisfactory fit.

Deleted: s 3 and

Formatted: Superscript

Deleted: green

Deleted: highlighted in yellow

Deleted: orange, or

Formatted: Superscript

Formatted: Superscript

Deleted: <#>**LOTUS Model Proxy Tests: Adding Tropopause Pressure (% difference in Std. Error of Model** ...

a)

a) LOTUS Model Test: Difference [%] in Standard Error: Tropopause Pressure vs Reference Model																
Pressure	Umkehr	sa/Davos			Haute-Provance			Boulder			Mauna Loa			Lauder		
(hPa)	Layer	UMK	COH	SND	UMK	COH	SND	UMK	COH	SND	UMK	COH	SND	UMK	COH	SND
1-2	9		0.3			0.1			0.5			1.4			3.0	
2-4	8	-0.7	-0.5		-0.1	-0.4		-0.2	0.4		-0.6	-0.3		1.3	2.6	
4-8	7	-0.3	0.0		0.3	1.3		0.3	-0.2		2.6	0.3		3.7	1.4	
8-16	6	-1.1	-0.7		0.0	0.3		0.7	-0.2		0.6	0.8		3.1	5.4	
16-32	5	-0.2	2.1	-0.9	1.1	5.3	2.4	-0.4	0.6	0.6	4.5	9.3	2.7	0.0	0.7	2.4
32-63	4	6.6		6.1	5.9		9.9	3.4		7.5	7.0		6.1	8.0		9.4
63-127	3	12.8		10.2	12.8		10.7	6.8		6.0	5.8		4.6	9.8		7.9

b)

LOTUS Model Test: Difference [%] in Standard Error: Equivalent Latitude vs Reference Model																
Pressure	Umkehr	Arosa/Davos			Haute-Provance			Boulder			Mauna Loa			Lauder		
(hPa)	Layer	UMK	COH	SND	UMK	COH	SND	UMK	COH	SND	UMK	COH	SND	UMK	COH	SND
1-2	9		8.4			2.9			1.9			-7.2			2.9	
2-4	8	-0.5	0.7		0.1	1.2		-0.4	1.5		-3.5	-5.4		1.0	3.1	
4-8	7	3.8	3.2		2.1	0.6		5.4	4.1		-2.6	-3.9		0.5	1.2	
8-16	6	6.1	8.3		2.5	10.9		2.4	7.8		5.3	7.8		3.4	7.7	
16-32	5	7.9	10.6	5.9	1.9	13.4	8.7	-1.9		1.4	0.3	0.7	0.7	0.8	3.9	-1.1
32-63	4	-1.4		-1.8	3.2		0.6	-0.2		-0.5	0.3		1.0	-0.2		-0.6
63-127	3	1.3		2.0	-1.4		-3.3	-0.8		-0.4	9.6		2.3	1.4		0.6

c)

c) LOTUS Model Test: Difference [%] in Standard Error: QBO C/D vs Reference Model																
Pressure	Umkehr	Arosa/Davos			Haute-Provance			Boulder			Mauna Loa			Lauder		
(hPa)	Layer	UMK	COH	SND	UMK	COH	SND	UMK	COH	SND	UMK	COH	SND	UMK	COH	SND
1-2	9		-1.6			-0.2			0.5			-0.5			-3.3	
2-4	8	-0.8	3.1		-0.3	9.1		2.9	4.6		-3.5	-0.3		-1.7	-0.4	
4-8	7	-0.1	1.5		0.3	3.3		-2.7	-1.2		-6.1	-4.2		1.8	1.4	
8-16	6	0.5	-1.3		1.1	-0.3		-2.4	0.7		-0.4	0.8		-2.5	-2.9	
16-32	5	-0.8	1.3	0.0	-0.9	3.0	2.8	0.6	0.6	0.1	7.1	8.4	10.1	-3.1	-0.6	1.1
32-63	4	0.9		0.2	2.0		-0.9	2.7		-1.8	2.9		6.5	-1.6		-1.9
63-127	3	-0.3		-0.8	5.7		-0.2	0.3		-4.2	0.4		3.0	-2.8		-3.2

d)

d) LOTUS Model Test: Difference [%] in Standard Error: AO/AAO vs Reference Model																
Pressure	Umkehr	Arosa/Davos			Haute-Provance			Boulder			Mauna Loa			Lauder		
(hPa)	Layer	UMK	COH	SND	UMK	COH	SND	UMK	COH	SND	UMK	COH	SND	UMK	COH	SND
1-2	9		1.2			-1.6			0.3			-1.9			-0.5	
2-4	8	-0.8	0.0		-3.8	-1.2		-0.8	-0.4		-2.1	-2.4		0.8	-1.9	
4-8	7	-0.7	1.7		-4.2	-2.6		3.2	4.7		1.2	-3.4		1.2	-1.2	
8-16	6	-0.2	-0.6		-2.4	-3.9		1.2	0.5		1.6	-1.6		-0.3	2.5	
16-32	5	-1.2	-0.4	-1.2	0.5	-2.1	-2.4	0.2	-0.6	-2.1	3.9	1.8	0.7	-1.6	-1.5	1.4
32-63	4	5.8		7.8	0.4		4.6	-1.2		-1.7	7.6		1.7	-0.7		2.5
63-127	3	13.1		12.9	5.5		6.8	-1.4		-3.2	4.4		1.3	-1.1		2.4

Formatted: English (United States)

Deleted: <#>LOTUS Model Proxy Tests: Adding Equivalent Latitude (% difference in Std. Error of Model) ...

Formatted: English (United States)

Deleted: <#>LOTUS Model Proxy Tests: Adding QBO CD (% difference in Std. Error of Model) ...

Formatted: English (United States)

Deleted: <#>LOTUS Model Proxy Tests: Adding AO/AAO (% difference in Std. Error of Model) ...

e)

e) LOTUS Model Test: Difference [%] in Standard Error: NAO vs Reference Model																
Pressure	Umkehr	Arosa/Davos			Haute-Provance			Boulder			Mauna Loa			Lauder		
(hPa)	Layer	UMK	COH	SND	UMK	COH	SND	UMK	COH	SND	UMK	COH	SND	UMK	COH	SND
1-2	9		0.5			-2.5			-0.2			-3.7			-1.3	
2-4	8	-0.2	0.0		-3.1	-1.9		-0.4	-1.9		-1.7	-3.8		-2.0	-1.8	
4-8	7	-0.6	0.7		-2.0	-2.0		0.0	3.9		2.6	-1.1		-2.8	-2.4	
8-16	6	0.2	-0.9		-1.7	-3.4		-2.2	-2.8		2.4	-0.4		-2.0	-0.5	
16-32	5	-0.5	-1.2	-1.1	0.7	-2.2	-4.0	-0.4	-1.4	-1.2	-1.4	-0.7	-3.0	-2.5	-4.3	-1.5
32-63	4	2.6		3.1	-0.6		-0.9	-0.2		0.4	1.5		-0.8	-2.6		-4.8
63-127	3	10.6		6.7	2.7		1.0	0.4		-2.7	1.7		-0.5	-2.3		-4.9

f)

LOTUS Model Test: Difference [%] in Standard Error: Eddy Heat Flux vs Reference Model																
Pressure	Umkehr	Arosa/Davos			Haute-Provance			Boulder			Mauna Loa			Lauder		
(hPa)	Layer	UMK	COH	SND	UMK	COH	SND	UMK	COH	SND	UMK	COH	SND	UMK	COH	SND
1-2	9		5.0			4.5			4.4			-3.2			0.2	
2-4	8	-1.4	4.6		2.6	6.0		3.1	8.8		-1.6	-3.3		0.7	1.9	
4-8	7	-2.7	-3.4		-0.4	-3.9		-3.0	-2.3		5.0	-4.4		1.8	4.5	
8-16	6	-3.1	-3.2		-2.5	-4.8		-2.4	-3.5		-1.1	0.4		-0.2	1.1	
16-32	5	-3.4	-2.8	-3.2	-2.2	-3.7	-2.5	-2.6	-2.4	-2.5	9.3	-0.4	4.3	0.7	2.4	0.7
32-63	4	-1.9		-1.6	-2.0		-1.8	-2.7		-3.5	8.8		3.1	1.9		1.6
63-127	3	1.5		1.4	-0.9		-1.6	-2.5		-3.8	0.9		0.9	2.1		1.1

Table 7: Change in Standard Error (SE) of the post-2000 trend estimate, in percent of SE of Reference Model for adding single predictors. Panel a: Tropopause Pressure; b: Equivalent Latitude; c: QBO terms C and D; d: AO/AAO; e: NAO; f: Eddy Heat Flux. **Cells with reduced (increased) SE have blue (red) background, while cells with low impact changes (<0.5 %) have no colours.**

5.1 Tropopause pressure (TP)

Adding the TP proxy to the standard LOTUS model produces the most consistent results between different techniques (COH, Umkehr and ozonesonde) and also have similar magnitude of standard error changes among different latitudes (i.e. Arosa/Davos, OHP, Boulder, MLO, Lauder). The most significant impact in improving the SE is found in the lower stratosphere (layers 3, 4) and in the middle stratosphere (layer 5) at the MLO tropical station. The impact of the TP proxy on the COH trend uncertainty in the model stratosphere (layer 5) is somewhat larger, likely due to the satellite AK extending into the lower stratosphere. Similarly, larger reduction of the standard error in the Umkehr trends in the lowermost stratosphere (layer 3) in comparison to the AK-smoothed ozonesonde record could be due to sampling biases in the ozonesonde record. Adding the TP proxy to the Reference Model improves the adjusted R^2 in layers 3–5, whereas the SE improvements are also consistent across geo-locations and measurement techniques. The TP proxy only explains ozone variability near the tropopause because changes in both parameters are linked to the same dynamical processes (i.e. irreversible mixing). In the middle and upper stratosphere ozone variability is not linked to the processes that change TP, thus using this proxy add error to the model fit. Several improvements resulted in adjusted R^2 to exceed the 0.3 threshold (Umkehr at OHP in layer 3, sonde and Umkehr at Lauder and MLO in layer 4) and in many cases the adjusted R^2 increased by more than 0.02.

Deleted: ¶
LOTUS Model Proxy Tests: Adding NAO (% difference in Std. Error of Model)

Formatted: English (United States)

Formatted: English (United States)

Formatted: English (United States)

Deleted: <#>LOTUS Model Proxy Tests: Adding Tropopause Pressure (Adjusted R^2 of Model)

Formatted: Superscript

Deleted: sonde at OHP in layer 4, and

Formatted: Superscript

Deleted: s 3 and

Formatted: Superscript

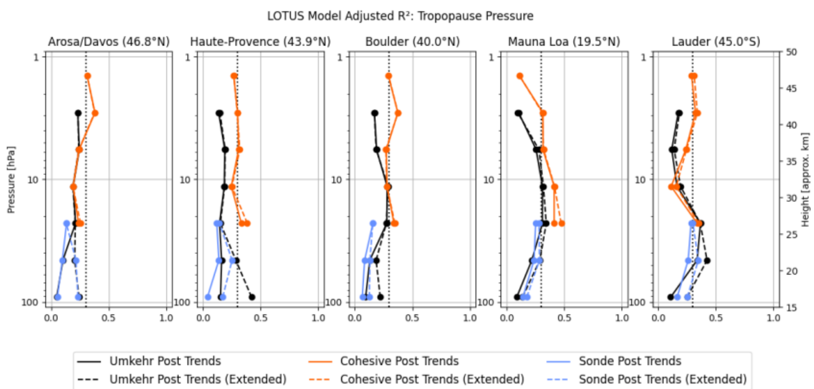


Figure X a) Similar to Figure 4, but adjusted R^2 results are shown for both Reference model (solid line) and the Extended model (Dashed line, Full) for COH overpass (orange), Umkehr (black) and ozonesonde (blue) trend. Extended model includes additional TP proxy.

5.2 Equivalent Latitude (EqLat)

In the mid-latitudes, the addition of EqLat as a predictor shows consistent results across measurement techniques and stations with few exceptions. The reduction in the SE of the model fit is evident in the COH data in the upper stratosphere (above 4 hPa or ~ 40 km), but is less pronounced in Umkehr profiles. The impact on MLO SE of the trend fit in the upper stratosphere is negative (in both COH and Umkehr records) which can be explained by the fact that the EqLat is much closer to geometric latitude near the equator than at the middle/high latitudes and therefore its u se as a proxy would not provide any additional information in interpretation of the tropical upper stratospheric ozone variability. It could also suggest that the addition of EqLat will overfit the record.

The ozone record trend fits in the middle stratosphere (32–4 hPa or 25–40 km) benefit from adding the EqLat proxy at most locations. Improvement in the SE of the trends in the lower stratosphere (127–63 hPa or ~15–20 km) is minimal, limited to some locations and instrumental records (Arosa/Davos Umkehr and HOH sonde, MLO Umkehr and sonde, and Lauder Umkehr and ozonesonde), which could be related to the location of subtropical jet that modulates mixing of tropical and subtropical (and occasionally polar) air masses and influences the strat/trop exchange. Unexpectedly, the addition of the EqLat proxy to the MLR statistical model for trend detection in Boulder Umkehr and ozonesonde lower stratospheric ozone records increases the uncertainties of the fit, while the influence of subtropical jet on Boulder lower stratosphere is well known (Manney et al, 2018). Perhaps, the data analyses also need to consider the tropopause variability.

In terms of the impact on the adjusted R^2, the EqLat proxy significantly improves model fit for multiple instruments, mostly in layers 5–7, and in COH fit in layer 9. The adjusted R^2 improvements also often exceeded 0.3 threshold. No significant improvement is found in the ozonesonde model fit in layer 5 with the exception of the OHP and

Hohenpeißenberg records (0.1 increase).

Deleted: 09

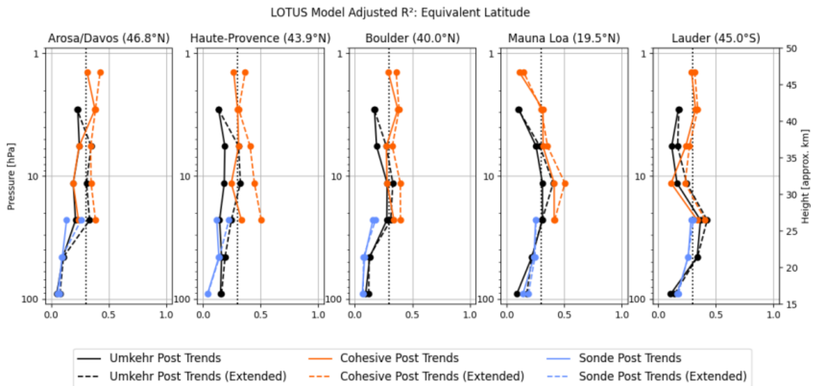


Figure X b) the same as a), but Extended model includes Equivalent Latitude proxy.

5.3 Extra QBO terms C and D

QBO is an important driver of ozone variability at tropical stations. Based on the results of adding 2 extra terms of the QBO to the standard model, the recommendation could be to exercise this option only for the tropical station trends. At the Northern middle latitudes (i.e in Arosa/Davos, OHP and, to a lesser degree, in Boulder) an improvement to the trend SE uncertainties in layer 8 is noted. There seems to be a similar pattern for the upper stratosphere in trends derived with Heat Flux. Tweedy et al. (2017) show that the first two EOFs of the QBO did not describe the anomalous QBO behavior, while Anstey et al. (2021) show that the addition of two more EOFs of the QBO could capture the effect of the disruptions on the zonal winds. Therefore, including additional QBO EOFS could benefit attribution of ozone variability in the middle stratosphere (layers 4 and 5) in the tropical latitudes (reduced errors in MLO/Hilo trends) and in the upper stratosphere (layer 8 in COH and in some Umkehr trends) in the NH middle latitude stations (Arosa/Davos, OHP, Boulder) related to the global circulation pattern that are also represented by the Heat Flux proxy. A slight reduction in the errors at SH middle latitude (sonde at Lauder, New Zealand) could be invoked by the EqLat variability that has a small correlation with the QBO-D proxy and sampling bias. Reduction of SE in the trend fit of the layer 5 ozonesonde (up to 2.8 %) and COH (up to 3.0%) records at OHP is not found in the Umkehr results, which suggests overfitting and sampling bias (see results in Appendix D).

Deleted: Umkehr and

Deleted: COH and

Deleted: (up to 2 %) is

Deleted: or COH

The addition of extra QBO terms slightly improves the adjusted R^2 model fit (see Figure X, c) for all COH station overpass records in layer 8 (except at MLO) and occasionally improves Umkehr adjusted R^2 (Boulder and Luader). The most significant improvement is found at MLO in layers 3–5 in all three instrument records.

Formatted: Superscript

Deleted: except at MLO

Formatted: Superscript

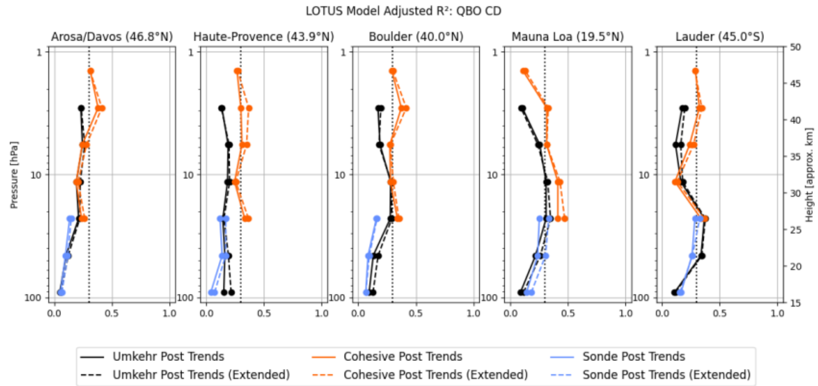


Figure X c) the same as a), but Extended model includes 2 extra QBO terms as an additional proxy.

Formatted: English (United States)

5.4 Arctic and Antarctic Oscillations (AO/AAO)

AO/AAO proxies reduce SE (blue colored cells) in the lower stratosphere (layers 3 and 4) at Arosa/Davos, OHP, and MLO, although the reduction somewhat differs between the Umkehr and ozonesonde records. At the same time, at Boulder and Lauder the SE does not show an improvement after the addition of the AO/AAO proxy (AAO is used instead of AO at Lauder). In the middle stratosphere (layer 7), a reduction in SE is found over Boulder in both COH and Umkehr records. The addition of AO/AAO proxies improves the SE of the trend at MLO and Lauder but only in Umkehr records, while it worsens the COH SE. At Lauder, the COH SE in layer 6 shows an improvement, but not in Umkehr record. Since results in the middle stratosphere (layers 5–7) are not always consistent among different techniques (reductions are not in the same layers) it could indicate statistical model overfit into the record's noise, or vertical smoothing of the Umkehr or COH technique that combines ozone variability in the layer with a portion of ozone variability in the adjacent layers, thus partially or completely reducing the correlation with the proxy.

Deleted: green

The addition of the AO predictor increases the adjusted R^2 in the lower stratosphere at Arosa/Davos, OHP and MLO. Also, a small enhancement of the adjusted R^2 is seen in the middle and upper stratosphere, including in Umkehr layers 6 and 7 and COH layers 6, 7 and 9 over Boulder, as well as in Umkehr fit in layers 5–7 at MLO, and at Lauder (AAO) for Umkehr and COH records in layer 6. These results are not very consistent across different geolocations, but seem to be consistent across instrumental records at some stations (Umkehr and ozonesonde in the lower layers, and COH and Umkehr in the upper layers).

Formatted: Superscript

Formatted: Superscript

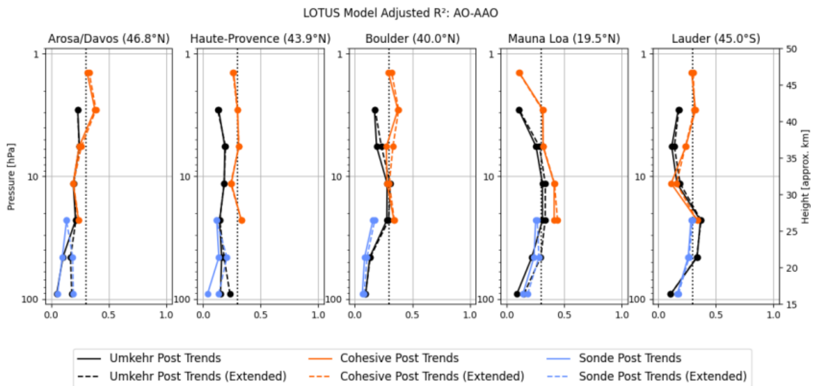


Figure X d) the same as a), but extended model includes AO/AAO proxy.

5.5 North Atlantic Oscillation (NAO)

Including the NAO proxy in the trend model appears to have a similar pattern (i.e., in latitude and altitude) of changes in the standard error as compared to the result of inclusion of the AO/AAO proxy. It is not a surprise, since indices of the NAO and AO are highly correlated in time due to their common link to the downward propagation of stratospheric anomalies. Standard errors are somewhat reduced in the lower stratospheric layers at the middle NH latitude and tropical Umkehr records, but the change is less significant than in AO/AAO cases. The impacts on ozonesonde trend uncertainties are very minimal and inconclusive at Boulder (layers 5 and 4), Hohenpeißenberg (layer 3 and 4), and OHP (layer 3). The impacts on Lauder are similar or stronger (SE is increased for both Umkehr and sonde records) to the impacts of the AO/AAO. In the middle and upper stratosphere, the standard errors are typically increased. The exception is found in layer 7 of the COH record at Boulder and Arosa/Davos, and in layers 6 and 7 of the Umkehr record at MLO. Similar negative results are found when AO/AAO proxies are added, which suggests that the observed time series are overfitted and potentially some instrumental or sampling anomalies are misinterpreted with addition of these proxies.

Formatted: English (United States)

Deleted: 3

Deleted: 5

Deleted: ,

Deleted: 4

Deleted: and MLO (layer 3) records

Deleted: reduced

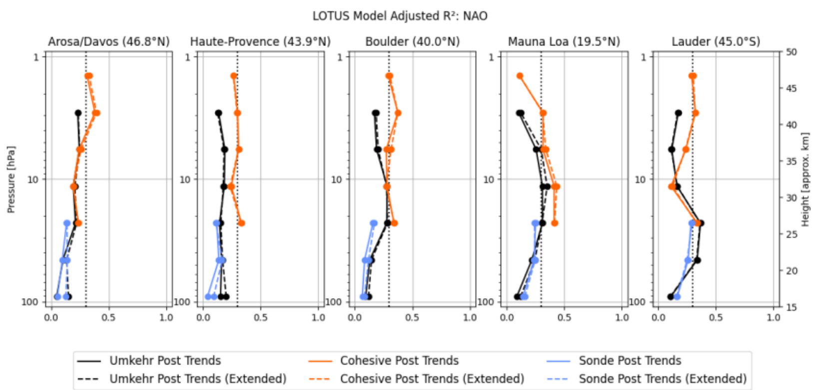


Figure X e) the same as a) but extended model includes NAO as an additional proxy.

5.6 Eddy Heat Flux (EHF)

The EHF represents a dynamical proxy for assessment of the impact of the Brewer Dobson Circulation (BDC). It is expected to have an impact on the upper stratospheric ozone by accelerating the transport in the upper branch that brings more ozone at higher latitudes (i.e. Arosa/Davos) and middle latitudes (i.e. OHP, Boulder, and Lauder). It could possibly represent changes in the lower branch of the BDC circulation and the expansion of the tropical band, thus modulating ozone in the lower stratosphere at tropics (i.e. MLO). In the Southern middle latitudes (i.e. Lauder), the correlations could be related to the shift in the subtropical wave activities to the higher latitudes in response to the ozone hole healing.

The addition of the EHF predictor leads to the reduced SE uncertainties in the upper stratosphere in COH and Umkehr trends at OHP and Boulder, and in COH only trends at Arosa/Davos. It has a much smaller reduction of SE for the Lauder trend and even an increase in uncertainties if used to fit upper stratospheric ozone time series at MLO. At the same time, the SE in the Umkehr and ozonesonde middle stratosphere (layers 4–5) at MLO is substantially reduced, including smaller improvements at Lauder. In the lower stratospheric (layer 3) ozone trend SE in Umkehr and sonde records at MLO, Lauder and Arosa/Davos are somewhat reduced when using the EHF proxy.

Addition of the EHF predictor seems to have an impact in the upper stratosphere increasing the adjusted R^2 for COH records in layers 8 and 9 in all but MLO or Lauder records, which indicates impact of the BDC upper branch on the middle NH latitudes. In contrast to the COH, the Umkehr adjusted R^2 has not changed significantly, which possibly suggests a high measurement noise in the station records. There is, however, a small increase in adjusted R^2 in the Umkehr record in layer 7 at MLO (whereas COH does not show a change).

The increase in adjusted R^2 is found at MLO in Umkehr and sonde layers 4 and 5, including a small increase in layer 3, which probably is related to the EHF-driven changes in the middle stratosphere . Ozone variability in Umkehr and sonde records at MLO appears to contain information about the circulation changes in the shallow BDC branch.

Formatted: English (United States)

Formatted: Superscript

Formatted: Superscript

Formatted: Superscript

Formatted: Superscript

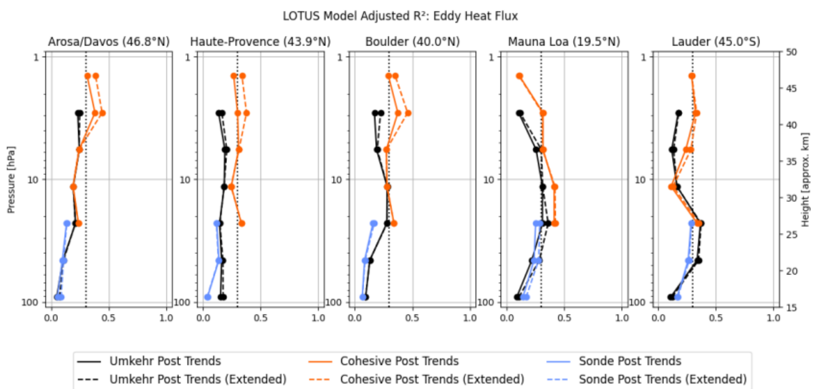


Figure X e) the same as a) but extended model includes EHF as an additional proxy

6 The Full Model - adding multiple predictors

In this paper we seek to develop an improved model and thus trend estimates for point located measurements of ozone through modifications of a model optimized for zonal data. Our criteria for model improvement are based on reduction of the SE of the trend with either improvement (at best) or moderate impact (at worst) on the model adjusted R^2. From the results of the previous section, we see several opportunities to improve the model and improve confidence in the trend estimates. This section examines if the gains of the above are improved while adding several predictors together. As stated above the TP as a predictor exhibits the most consistent results for all stations and measurement techniques. The other predictors have successes in SE reduction, but only at some layers, and some stations. Some results are instrument dependent.

Based on the tests above we expect combining predictors can improve the model fit and trend SE reduction, but it is clear that the predictor selection should vary by station and level. Appendix E details the choices made for the Full Model which combines 1 to 3 additional proxies beyond the Reference Model.

6.1 Predictors added for the Full Model

Reduction of the SE of the trend while improving (or at least not impacting) model adjusted R^2 is the basis of predictor choice for the Full Model. To qualify a predictor should exhibit consistent results for all measurement techniques. Improvement at multiple stations is preferred to single station improvements. In general, we avoid combining highly correlated predictors. Table 9 shows final choices for the Full Model.

LOTUS Full Model predictor selection					
	Arosa/Davos	**OHP**	**Boulder**	**MLO**	**Lauder**
Layer					

Formatted: English (United States)
Formatted: Superscript
Formatted: Superscript

9	EqLat	EqLat	EqLat	Reference only	EqLat
8	EqLat	EqLat	EqLat	Reference only	EqLat
7	EqLat	EqLat	EqLat	Reference only	EqLat
6	EqLat	EqLat	EqLat	EqLat	EqLat
5	EqLat	EqLat	EqLat	EqLat, QBO CD, AO	EqLat
4	TP, AO	TP, AO	TP	TP, QBO CD, AO	TP
3	TP, AO	TP, AO	TP	TP, QBO CD, AO	TP

Table 9: Added predictors for the Full model are tuned for each layer and each station. For layers 7 to 9 the SE and adjusted R^2 parameters at MLO are not improved by additional predictors, and the original LOTUS based Reference Model is used. Appendix E explains the logic of the predictor selection.

6.2 Impact of the Full Model on trends

Figure 5a shows the trends for the stations (with COH overpass) for the Reference and Full Models. An impact of the Full Model on ozone trends derived in the upper stratosphere (above 16 hPa) is neutral. Addition of proxies to the LOTUS model does not change trends which remain the same magnitude as those derived using the Reference Model, i.e., positive and statistically significant at the SH and MH middle latitudes and over tropics. The largest difference (outside of the SE uncertainty) between upper stratospheric Umkehr and COH trends is found over Boulder, MLO and Lauder.

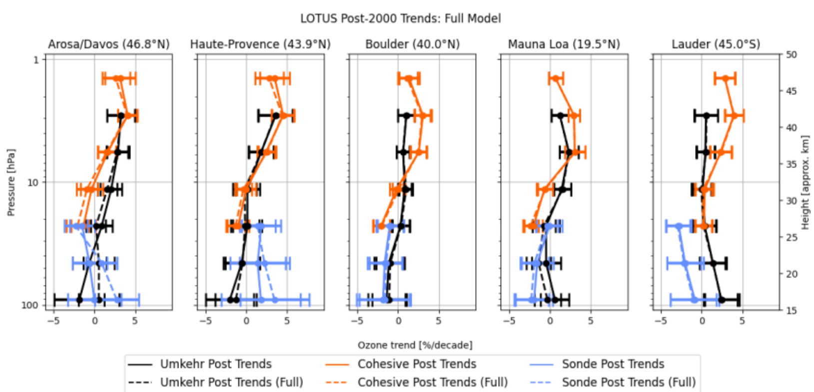


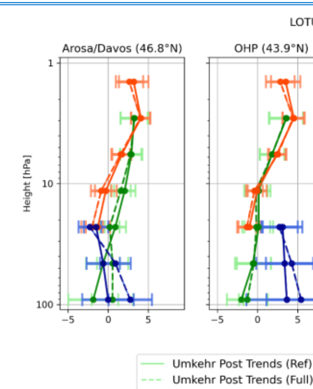
Figure 5a: Post 2000 trends for the Full and Reference Model. In this figure the COH data shown in orange is the overpass data. Solid lines depict Reference Model values (unchanged from Fig. 3). Dashed lines depict Full Model values for all 3 instrument types.

Formatted: Superscript

Deleted: ¶

Deleted: i.e.

Formatted: Heading 2



Deleted:

Formatted: English (United States)

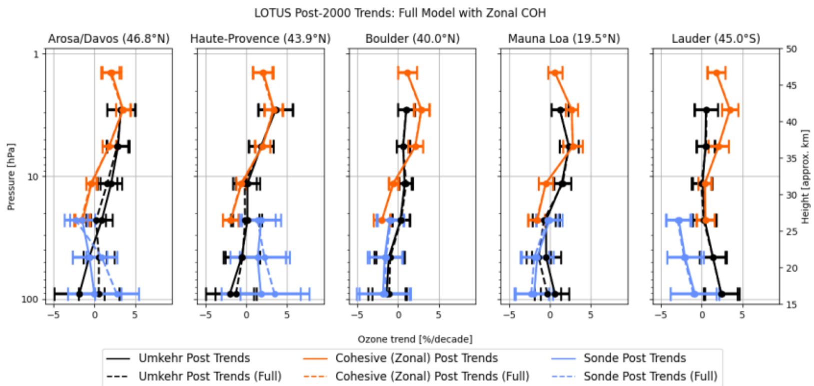


Figure 5b: Post 2000 trends for the Full and Reference Model. In this case the orange lines are with the zonal data instead of the COH overpass data. Dashed lines depict Full Model values for all 3 instrument types. The Umkehr and sonde trends are unchanged from Fig. 5a.

In the middle stratosphere, additional proxies do not change trend values across locations and instrumental records (outside of the SE). At OHP, Boulder and Lauder Umkehr trends in layer 6 (8–16 hPa) are barely positive while COH trends are negative. At Arosa/Davos and MLO, COH trends in layer 6 are barely negative and Umkehr trends are significantly positive. Most COH trends in layer 5 (16–32 hPa) are statistically negative (except at Lauder), while Umkehr trends are near zero.

In the lower stratosphere, Umkehr and sonde trends Arosa/Davos and MLO change after the Full model is used. However, Umkehr and sonde trend changes at MLO are within the SE and therefore can be deemed not significant. Ozonesonde trends at Arosa/Davos in layer 3 (125–63 hPa) change from zero to positive. Umkehr trends at Arosa/Davos in layer 3 change from negative to near zero. Large differences between ozonesonde and Umkehr trends at Lauder and OHP remain unchanged after the Full model is applied although respective SE envelopes overlap.

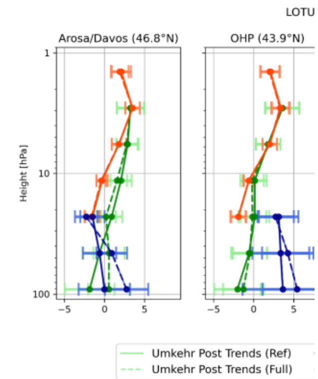
Figure 5b also shows the trends for the Reference and Full Models, but the COH data shown is the associated zonal data relevant to each station. Incorporation of the additional proxies does not change the trend values for the zonal COH data. Impact on error estimates for the trends are discussed next.

6.2 Impact of the Full Model on the Trend SE

Table 10 summarizes the reduction in the SE for the Full model. Selection of the EqLat predictor for the Full model in the layers 5–9 and for all stations (except MLO/Hilo, to be discussed later) shows the improvement in the SE (as discussed in the previous section). Also, the TP predictor is selected for inclusion to the Full model for trend analyses at Boulder and Lauder stations in layers 3 and 4. The combination of several predictors are used for individual stations based on the additional reduction in the SE. For the Arosa/Davos and OHP stations we select a combination of the TP and AO to reduce the SE almost twice as much in some layers. Inclusion of AO proxy is in support of the interpretation of seasonal and interannual ozone variability recorded over stations in Europe that are north of 40 degrees latitudes and are exposed to the seasonal events of ozone depleted air masses transported from the Polar region during the

Formatted: Caption, Indent: Left: 0", First line: 0", Widow/Orphan control

Deleted:



Formatted: English (United States)

spring season (Steinbrecht et al., 2011; Manney et al., 2011; Knudsen and Grooss, 2000; Fioletov and Shepherd, 2003; Zhang et al., 2017; Weber et al., 2022a). The strong impact of AO/AAO on the lower stratosphere ozone variability are not detected in Boulder or Lauder and we choose not to include it in the Full model for trend analyses at these stations.

LOTUS Model Test: Difference [%] in Standard Error: Full Model vs Reference Model																	
Pressure	Umkehr	Arosa/Davos			Haute-Provance			Boulder			Mauna Loa			Lauder			
(hPa)	Layer	UMK	COH	SND	UMK	COH	SND	UMK	COH	SND	UMK	COH	SND	UMK	COH	SND	
1-2	9		8.4			2.9			1.9			0.0			2.9		
2-4	8	-0.5	0.7		0.1	1.2		-0.4	1.5		0.0	0.0		1.0	3.1		
4-8	7	3.8	3.2		2.1	0.6		5.4	4.1		0.0	0.0		0.5	1.2		
8-16	6	6.1	8.3		2.5	10.9		2.4	7.8		5.3	7.8		3.4	7.7		
16-32	5	7.9	10.6	5.9	1.9	13.4	8.7	-1.9		1.4	13.6	13.0	13.3	0.8	3.9	-1.1	
32-63	4	8.7		10.0	6.1		9.4	3.4		2.7	17.3		10.3	8.0		7.4	
63-127	3	20.3		18.5	13.5		12.8	6.8		2.2	8.0		5.6	9.8		6.8	

Table 10: Change in post 2000 trend SE in the Full Model as a % difference of the Reference Model. Color coding is the same as introduced in Table 7.

The MLO/Hilo location is close to the Tropical belt and therefore has different processes impacting stratospheric ozone variability as discussed in the previous section. We find that EqLat proxy can be added to the Full model in layer 6 and 5 (similar to other stations); however, above layer 6, EqLat or TP is not useful for interpretation of tropical ozone variability and therefore we believe the trend model in these layers should remain as it currently is used in Godin-Beekmann et al. (2022) analyses. The EqLat and TP are mildly correlated (-0.4) in the stratosphere, and therefore we decided against combining both of these proxies in the Full model. However, we also found that adding AO and QBO C/D proxies in layers 3, 4 and 5 improved the model fit and reduced the SE. These combined additional proxies are not correlated and reduce SE more than when using them separately.

The Full Model showed impacts on the SE in the upper stratosphere (above 8 hPa). The trend errors were reduced with the exception of Umkehr trends at 4–2 hPa over Boulder and Arosa/Davos where errors did not change. No changes in SE are found at MLO with additional proxies, thus the Full Model is kept the same as the Reference Model for this station in the upper stratosphere.

Similarly, in the middle stratosphere SE were mostly reduced after the Full Model was applied (except for slightly larger SE in trends derived from ozonesonde at OHP and from Umkehr at Boulder).

After applying the Full Model in the lower stratosphere, we still found high uncertainty due to higher ozone variability (natural variability), but SE were reduced. Arosa/Davos and MLO Umkehr and sonde trends changed after Full Model was used. Change in ozonesonde trends at HOH in layer 3 (125–63 hPa) goes from zero to positive and trend detection becomes highly confident (p-value <0.05). Umkehr trends at Arosa/Davos in layer 3 changed from negative to near zero but results have low certainty (p-value >0.1). Larger trend differences remain between ozonesonde and Umkehr at Lauder and OHP after the Full Model is applied.

LOTUS Model Proxy Tests: (% difference of SE of Trend): overpass and zonal COH											
Height	**Umkehr**	**Arosa/Davos**		**OHP**		**Boulder**		**MLO**		**Lauder**	
(hPa)	**Layer**	**Overpass**	**Zonal**	**Overpass**	**Zonal**	**Overpass**	**Zonal**	**Overpass**	**Zonal**	**Overpass**	**Zonal**

Deleted: **LOTUS Model Proxy Tests: (% Difference in Std. Error of Model)** ...

Formatted: English (United States)

Deleted: ...

1-2	**9**	7.61	2.89	2.20	1.30	1.61	1.34	NA	NA	3.17	1.97
2-4	**8**	0.00	0.90	1.47	1.26	0.00	0.63	NA	NA	1.75	2.76
4-8	**7**	3.39	0.47	1.85	2.55	5.77	1.53	NA	NA	0.00	1.11
8-16	**6**	7.35	2.75	10.17	8.98	9.30	4.30	0.00	1.79	8.93	5.34
16-32	**5**	10.67	1.74	11.76	5.54	0.00	2.36	12.73	4.81	3.70	-1.11

Table 11: Change in Standard Error of Trend, as percent of Reference Model SE, for the COH overpass data and zonal data at the 5 ground stations. MLO Full Model in layers 9-7 is the same as the Reference Model (change is marked as NA).

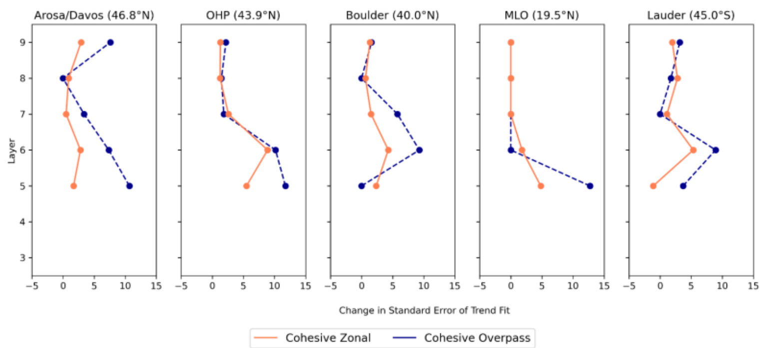


Figure 6: Change in Standard Error of Trend, as percent of Reference Model SE, for the COH overpass data (blue) and COH zonal data (red) at the 5 ground stations.

It is instructive to ponder if the addition of proxies that yield improvements via reduction of the standard error in the localized GB or overpass measurements also have the potential to improve uncertainties in the zonal data. To explore this Table 11 and Fig. 6 show the percent change in SE of the trend when adding the proxies for the Full model. Values are shown for both the COH overpass and the COH zonal data. In general, except when the improvement in the SE for the overpass COH is small (3% or less), addition of proxies has much less impact on the zonal results than on overpass results. This suggests that indeed the Reference LOTUS model is well tuned for zonal datasets, but can be improved with select addition of proxies for overpass or localized GB data.

6.3 Impact of the Full Model on adjusted R^2

Table 12 shows the adjusted R^2 for the Full Model. In the upper stratosphere, the Full Model increases the adjusted R^2 above 8 hPa (except in Umkehr at 4–2 hPa). Over MLO there is no change because the Full Model is kept the same as the Reference Model for layers 7, 8 and 9.

LOTUS Model Proxy Tests: (Adjusted R^2 of the Full Model)																
Height	**Umkehr**	**Arosa/Davos**			**OHP**			**Boulder**			**MLO**			**Lauder**		
(hPa)	**Layer**	**UMK**	**COH**	**SND**	**UMK**	**COH**	**SND**	**UMK**	**COH**	**SND**	**UMK**	**COH**	**SND**	**UMK**	**COH**	**SND**
1-2	**9**		**0.42**			**0.37**			**0.36**			0.11			**0.32**	

Formatted: Superscript
Formatted: Superscript
Formatted: Superscript
Formatted: Superscript

2-4	**8**	0.23	**0.39**		0.14	**0.31**		0.17	**0.39**		0.11	**0.32**		0.18	**0.34**	
4-8	**7**	**0.35**	**0.35**		**0.31**	**0.41**		0.27	**0.33**		0.26	**0.32**		0.17	0.27	
8-16	**6**	**0.31**	**0.35**		**0.33**	**0.45**		**0.33**	**0.40**		**0.40**	**0.51**		0.25	0.23	
16-32	**5**	**0.34**	**0.38**	0.26	0.25	**0.51**	0.23	**0.31**	**0.40**	0.18	**0.44**	**0.53**	**0.39**	**0.42**	**0.41**	0.29
32-63	**4**	0.23		0.25	0.29		0.27	0.19		0.18	**0.42**		**0.38**	**0.42**		**0.31**
63-127	**3**	**0.31**		**0.31**	**0.44**		0.21	0.22		0.11	0.19		0.24	0.25		0.21

Table 12: Adjusted R^2 of the Full Model. Values of 0.30 and above are indicated in Bold as a threshold to indicate a satisfactory fit. Compare to Table 4 containing values for the Reference Model.

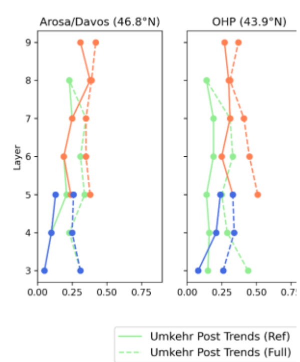
In the middle stratosphere (32–8 hPa) adjusted R^2 increases are found in all records (although smaller increases are found in ozonesonde and Umkehr records at OHP, Boulder and Lauder at 32–64 hPa). At Arosa/Davos, Boulder and Lauder the adjusted R^2 in the COH and Umkehr trend models increase and continue to be very close in value. The COH adjusted R^2 is larger at OHP and MLO than in Umkehr and sonde records thus suggesting that overpass conditions might have smoothed some natural variability observed in the GB records. In general, the adjusted R^2 is the largest at the 32–64 hPa level. This suggests that the Full Model shows an improvement for regional trend analyses in the middle stratosphere.

Although Umkehr and sonde trend changes at MLO in the lower stratosphere are within the SE and therefore can be deemed not significant, the adjusted R^2 is increased which suggests a better model fit in the Full Model. The adjusted R^2 increases in both Umkehr and ozonesonde data, while the largest increases are found in the Arosa/Davos, OHP and MLO records.

In the lower stratosphere, the adjusted R^2 remains low in both Umkehr and sonde records at Boulder (only TP is added for the Full model). While the p-values at 63–32 hPa are significantly reduced (see discussion in the next section), they still remain relatively high. These results suggest that additional research is needed to identify the best set of proxies for Boulder records in the lower stratosphere. At Lauder, the ozonesonde record shows smaller adjusted R^2 as compared to Umkehr partially due to low sampling biases.

It is valuable to further explore the impact of the Full Model on the adjusted R^2 for the zonal and overpass COH data. Fig. 7a shows the adjusted R^2 for the Reference and Full Models at each of the 5 stations using the COH overpass data. In all cases the Full Model improves the adjusted R^2 except for MLO layers 7, 8 and 9 where the Full and Reference Model are identical. The most significant improvements are seen by Umkehr at layers 3 to 7, COH overpass at Layers 5, 6 and 7, and sonde layers 3–5. Figure 7b shows similar results using COH zonal data instead of overpass. There is practically no further improvement in the adjusted R^2 for the zonally averaged COH results (except for a small increase for MLO layer 5). Comparison of results reveals that for OHP the implementation of the Full model for the COH overpass data (Fig. 7a, dashed line) improves the adjusted R^2 to values nearing that of the Reference Model zonal data in layer 7 and below (Fig. 7b, solid line). For MLO and Lauder the use of the Full Model on the COH overpass data improves the adjusted R^2 over the Reference Model beyond the improvement seen in the COH zonal results for layers 5 and 6. At Arosa/Davos and Boulder the implementation of the Full Model does not fully reach the magnitude of the COH zonal adjusted R^2.

Deleted: 3
Deleted: 34
Deleted: 6
Formatted: Superscript
Formatted: Superscript
Formatted: Superscript
Formatted: Superscript
Formatted: Superscript
Formatted: Superscript
Formatted: Superscript
Formatted: Superscript
Formatted: Superscript
Formatted: Superscript
Formatted: Superscript
Formatted: Superscript
Formatted: Superscript
Formatted: Superscript
Formatted: Superscript



Deleted:

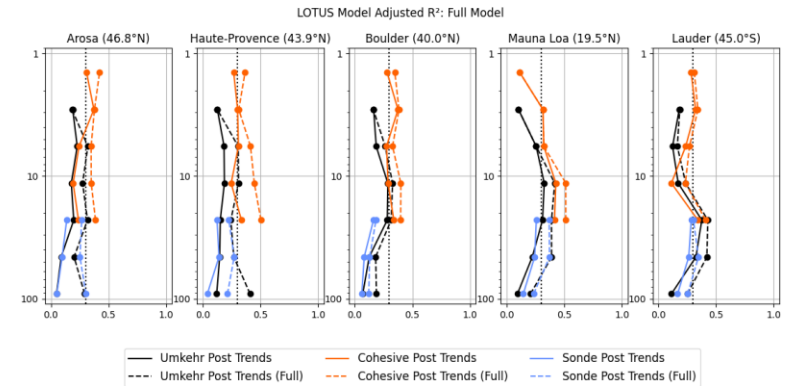


Figure 7a: Adjusted R^2 for the Full Model (dashed lines) and Reference Model (solid lines) at 5 stations. The COH data in this figure is the overpass data at each station.

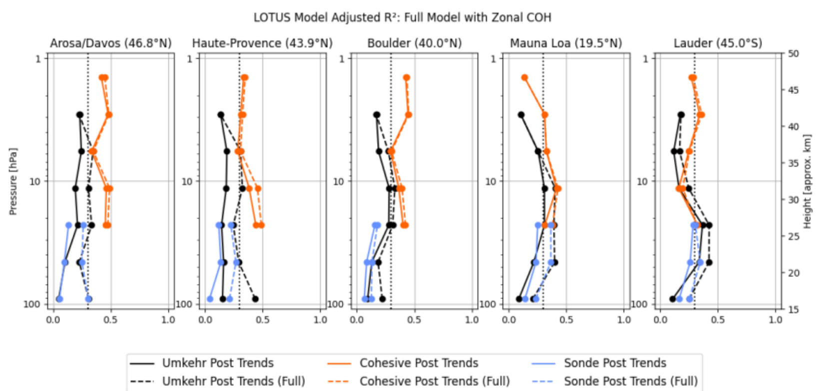
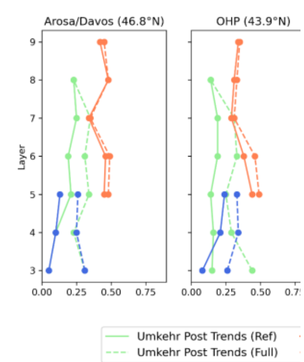


Figure 7b: Adjusted R^2 for the Full Model (dashed lines) and Reference Model (solid lines) at 5 stations. The COH data in this figure is the zonal data for each station. The Umkehr and sonde lines are identical to those in Fig. 7a.

Formatted: Superscript

Formatted: English (United States)

Formatted: Caption, Indent: Left: 0", First line: 0", Widow/Orphan control



Deleted:

Formatted: Superscript

Formatted: Indent: First line: 0"

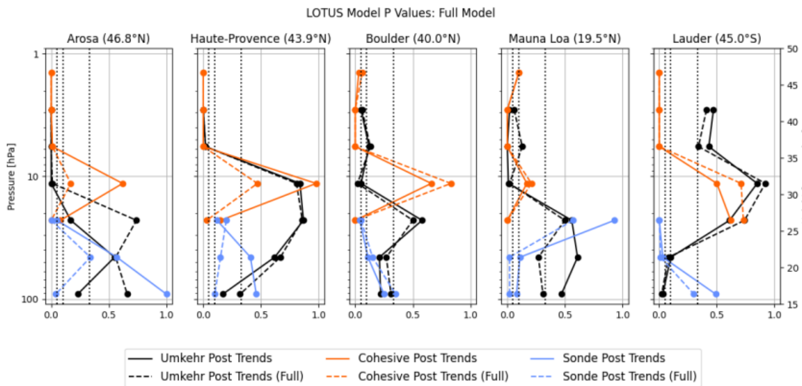


Figure 7c: the same as b, but for the p-values. Vertical dotted lines indicate limits for the high (<0.05), medium-high (between 0.05 and 0.1), medium (between 0.1 and 0.3) and low confidence (>0.3).

6.3 Examination of the p-values of the Full Model

In the upper stratosphere (above 8 hPa), the confidence in Umkehr trends remained high (see Figure 7c) for most stations except at Boulder (medium to low) and Lauder (very low, although some improvement was found). COH trends confidence was very slightly degraded over Boulder at 1–2 hPa, but mostly has not changed.

In the middle stratosphere (between 32 and 8 hPa), p-values were significantly reduced in COH records. At 8–16 hPa remained high, but at 16–32 hPa the confidence improved (continued) to high over Arosa/Davos and OHP (Boulder and MLO). In case of Umkehr analyses in layer 8–16 hPa at Arosa/Davos, Boulder and MLO the confidence remained high. However, at 16–32 hPa the Umkehr trend detection confidence was degraded over Arosa/Davos and Lauder. For the ozonesonde record, the p-values remained low (<0.05) except at MLO where some improvement was found after the Full Model was used, but the p-value remained high. It suggests that some instrumental records have either high atmospheric or instrumental noise and therefore perhaps high certainty in trend detection cannot be achieved with linear trend models. For near zero trends with high variability, the p-values are not a good criterion for trend detectability.

In the lower stratosphere (between 125 and 32 hPa), analyses of p-values for the Full Model fit show significant improvement for Umkehr trends at MLO between 63–32 hPa (while the p-value was increased at other stations at this level). In addition, improvement in p-values was found for ozonesonde trends at all stations. Specifically, very low p-values for the Full model were reached at Arosa/Davos (125–63 hPa), OHP (125–63 and 63–32 hPa), MLO (125–63 and 63–32 hPa), and Lauder (63–31 hPa).

Formatted: English (United States)

Formatted: Normal

Deleted: Table 13

Deleted: **LOTUS Model Proxy Tests: (P Value of Model)** ...

7 Summary of the Full Model findings.

We find that upper stratospheric trends in COH overpass and Umkehr records detect ozone recovery with high confidence ($p<0.05$) above 8 hPa (with the exception of near-zero positive Umkehr trends over Lauder and Boulder). We note the largest difference between Umkehr and COH trends (outside of the SE uncertainty) at Boulder, Mauna Loa and Lauder.

Confidence for the middle stratosphere (between 32 and 8 hPa) trends vary between high, medium and low. Although most of the trends are narrowly different from zero (especially when error bars are considered), there are some differences in results across instrumental groups: trends in COH and sonde (except at OHP) between 32 and 16 hPa tend to be small negative, while Umkehr trends are slightly positive. Some trends are statistically different from zero. However, instrument-specific error bars often overlap and thus making differences in trends not significant.

Confidence in lower stratosphere trends is highly variable and even lower than in the middle stratosphere due to higher ozone variability unaccounted for by Solar, QBO and ENSO proxies used in the Reference Model. However, high confidence ($p<0.05$) is still found in ozonesonde trends at Arosa/Davos, OHP, MLO and Lauder (although not at all layers). Umkehr trends in the lower stratosphere show lower confidence than ozonesonde trends (except at Lauder and Arosa/Davos in the lowermost altitudes). The low confidence levels could be related to the near-zero trends derived from Umkehr data, whereas ozonesonde trends are often different from zero lines. Also, we apply AK-smoothing to the sondes to account for the wide AKs in the Umkehr retrieval. We tested the impacts of the AK on ozonesonde trends (see Appendix A) and did not find any significant impacts. Most notably, ozonesonde and Umkehr trends significantly disagree in the lower stratosphere at OHP and Lauder and therefore require further investigation. The instrumental drifts and differences in Lauder trends are also discussed in Bjorkland et al. (2023 preprint) and are consistent with our findings.

8 Conclusions.

This paper is a follow up to Godin-Beekmann et al. (2022) with a focus on the GB record trend assessment. Therefore, our trend analyses focus on the questions:

1) Do proxies for evaluating trends of GB stations need to be different from those of the optimized set for zonal data?
2) Are station records representative of the small geophysical region or semi-global changes?
3) Do uncertainties of the zonal averaged trends improve with additional proxies?

The Full Model developed in this paper for station and overpass data adds proxies to the LOTUS models of Godin-Beckmann (2022). Our trend analysis of stratospheric ozone records from the Umkehr, ozonesonde and COH station overpass data at 5 geographical regions using the Full Model (LOTUS v 0.8.0) show similar trends to those published in Godin-Beekmann et al. (2022) paper. We analyze trends for instrumental records converted to 7 Umkehr layers that represent ozone changes in the upper, middle and lower stratosphere over NH and SH middle latitudes and over high tropics of the NH. We also analyze GB station records at Arosa/Davos, Hohenpeißenberg and OHP separately in contrast to the "European regional" trend analyses presented in Godin-Beekmann et al. (2022) and included COH overpass records for comparisons with the GB records. Our analyses include evaluation of the adjusted R^2 (aka goodness of the model fit), standard error and p-values.

We also investigate differences between satellite trends as detected in the records sampled for individual geographical locations (spatial and temporal overpass criteria) versus zonal average datasets. We find that COH overpass ozone records capture ozone variability of the ground-based station records (Umkehr and sonde) better than COH zonal data. We do not find that the COH zonal record is improved by using EqLat instead of geometric latitude to construct the dataset (see Appendix C), but EqLat can be an important additional proxy at some levels for GB data. To determine the improvement to the model fit we use the Standard Error and adjusted R^2 for the Full and Reference model fit. Using the Reference model for the zonal mean COH data we find slightly better adjusted R^2 than for the COH overpass data fit over the Northern middle latitude stations. This is expected as much of the variability of the overpass time series is reduced in the zonal average data. Therefore, we also explore the impact of additional predictors in the trend model fit applied to the more variable GB and satellite COH overpass data to determine if that will reduce the SE and improve the adjusted R^2. We also apply the Full model to the zonally averaged data to assess the benefits of additional proxies to further reduce trend uncertainties.

We find that adding predictors (with few exceptions) does not change the trends but often reduces SEs and increases the adjusted R^2 (with the exception of the upper stratospheric ozone trends at MLO). We also find that the p-values are useful for interpretation of improvements of the model fit in the data, although improvements in the SE do not always result in improved confidence in derived trends, especially when the trends are close to zero. In these cases we conclude that either longer records are needed to discern trend information outside of the atmospheric noise or further research into the inconsistencies between instrumental records and homogenization procedures is required. We also find the small changes in trends in the lower stratosphere and improvements in the model fit after additional proxies are used. However, the sampling tests indicate that trends can depend on the temporal selection of the records when AK are used to smooth ozonesonde high resolution profiles (see discussion in Appendix D).

This paper concludes that additional proxies bring improvements to trend detectability for GB and gridded satellite data analyses and better agreement is achieved between satellite overpass and GB trends. We also find that zonally averaged and gridded satellite records produce comparable trends over the studied middle latitudes and subtropical regions. Therefore, the GB trends are representative of the stratospheric ozone changes over the semi-global area. Finally, zonally averaged data do not benefit from addition of proxies beyond what LOTUS model uses for global trend detection whereas the uncertainties in GB and gridded trends are significantly reduced and sometimes (Boulder, MLO, Lauder) become comparable to the uncertainties of the zonally averaged trends in the upper and middle stratosphere. Based on analyses presented in this paper we strongly recommend using additional proxies for trend analyses of GB and gridded satellite stratospheric ozone records. Additional proxies should be selected based on the latitude and altitude of the observational ozone record to adequately represent stratospheric transport and mixing processes impacting interannual and seasonal ozone variability.

Formatted: Superscript

Formatted: Superscript

Formatted: Superscript

Formatted: Superscript

Formatted: Superscript

Deleted: in case of

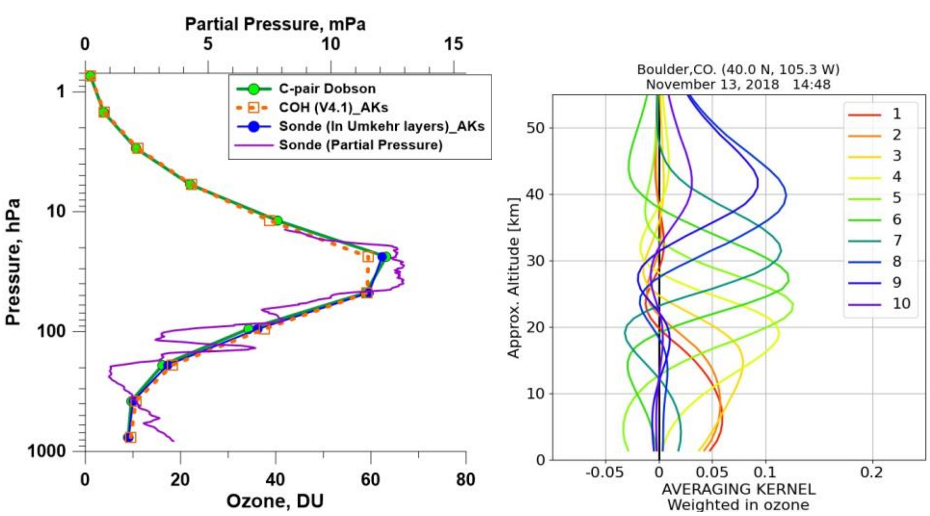
Appendices

Appendix A: AK Smoothing for ozonesondes

Ozonesonde profiles have high vertical resolution (purple line in Fig. A1) in comparison to the Umkehr (green solid line) or COH (orange dashed line) ozone profiles. Each Umkehr layer is referenced to the atmospheric pressure at the bottom of the layer, which is constructed using half of the pressure in the layer below. Averaging Kernels (AK) as shown in Fig. A1, panel b, define the granularity of the Umkehr vertical grid. In order to compare trends from three instrumental records in the same vertical system, we convert the ozonesonde and COH profiles to the Umkehr layers and DU. The COH overpass data is in units of DU, but on different layers than the defined Umkehr layers, so only vertical grid modification is required. The sonde profiles (purple thin line) are in units of partial pressure and are first converted to DU, then converted to the Umkehr grid (blue solid line in panel a). Conversion to the Umkehr grid can be done either by interpolation, or by AK smoothing. The equation describing the process of applying AK smoothing is

$$Ozone_{smoothed}(i) = Ozone_{apriori}(i) + \sum_j AK_{ij} \{Ozone_{true}(j) - Ozone_{apriori}(j)\}$$

where AK is the Averaging Kernel for layer i, $Ozone_{smoothed}$ is the smoothed ozone result, $Ozone_{true}$ is the ozonesonde profile, and $Ozone_{apriori}$ is the Umkehr a priori (climatological) profile. The AK for each Umkehr layer is used as a weighting function applied to the ozonesonde profile ($Ozone_{true}$) prior to the integration which simulates the Umkehr optimal estimation method used for estimating the ozone content in the targeted layer (Rodgers, 2000).



Deleted: $Ozone_{smoothed}(i) = \sum_j \{AK_{i,j} * Ozone_{tru}$

Deleted:

Figure A1: a) An example of ozone observations over the Boulder, CO station. The purple line is 100-m averaged ozone partial pressure (hPa) vertical profile measured by sonde on 13 November, 2018 . The green line with solid circles is the ozone profile derived from Dobson Umkehr observations on the same day. The blue line with blue dots is the ozonesonde profile converted to the Umkehr layers and smoothed with the Umkehr AK. The orange dashed line with open squares is the COH ozone profile observed over Boulder on the same day and interpolated to the Umkehr layer vertical grid. b) The Umkehr AK for the ozone profile derived from observations in Boulder on 13 November, 2018. Each line represents the smoothing function for one of 10 Umkehr layers (see color legend).

Although the ozonesonde measurement typically reaches altitudes between 32 and 10 hPa, the balloon often bursts before reaching the top of layer 6 (16 hPa), therefore only partially covering the ozone content in that layer. We also note that Umkehr AKs are relatively wide and therefore will incorporate (weight in) ozone variability from the layer above and layer below of the targeted Umkehr layer. (See layer 6, green line in Fig. A1, panel b.) Therefore, there are two sources of error in ozonesonde comparisons with Umkehr ozone in layer 6: a) burst level for ozonesonde does not reach the top of the layer 6, thus the integrated ozone is smaller than expected. b) the Umkehr AK for layer 6 is relatively wide and therefore the Umkehr layer partially contains information from above the burst altitude of the ozonesonde, thus making smoothed ozonesonde concentration lower than expected. In order to avoid these errors, we only show ozonesonde results up to layer 5.

Similarly, we explored smoothing COH profiles with Umkehr AKs. Figure A2 demonstrates the time series of the COH ozone over the Mauna Loa station. The trend model was fitted to the COH record with and without AK applied. The reference trend model included proxies and trends. To focus on ozone variability that contributes to the trends we subtracted the modeled ozone variability from the COH data and then added the trend component back. The COH record residuals in Fig. A2 are shown in Umkehr layers where COH is either smoothed with AK (red lines) or not (green lines). We notice that the AK-smoothing of the COH profile in layer 9 does not have a lot of independent information from layer 8. In this example it clearly shows that the trends in layer 8 are embedded in the COH layer 9 ozone time series, which was confirmed when we compared trends derived from the AK-smoothed COH in layers 8 and 9. In case of the integrated COH ozone record, the trends in layers 8 and 9 differed. In order to avoid biasing the COH trends at layer 9 we decided to not apply Umkehr AKs for COH smoothing and only use COH profiles interpolated into the Umkehr layers. This result makes sense since COH overpass data are derived from UV backscatter radiances also using an Optimal Estimation technique. COH overpass data has a comparable vertical resolution to Umkehr, simply with different layer definitions. Interpolation makes the most sense for rendering COH data in the Umkehr vertical coordinate system.

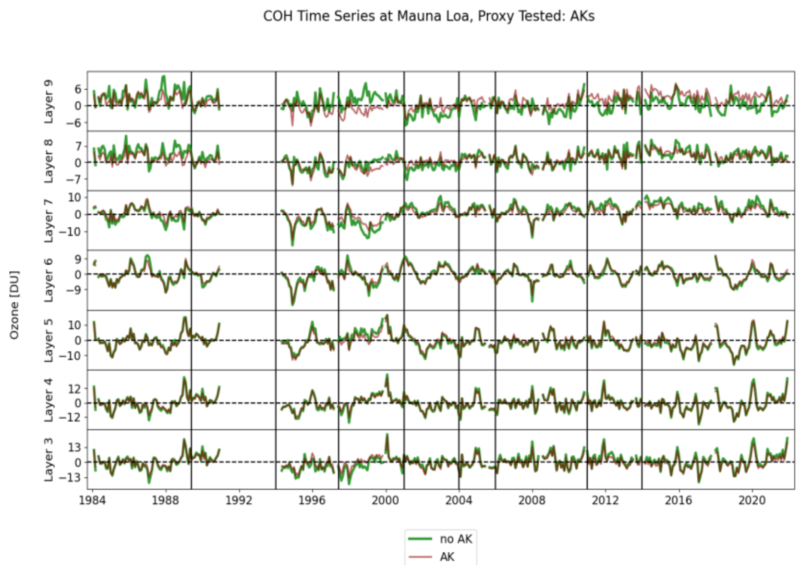


Figure A2: Modified residuals (seasonal cycle, Solar, QBO, and ENSO are removed, but trend is retained) of COH overpass data at Mauna Loa (20N, 156W). Red: AK smoothed to Umkehr layers; Green: Interpolated to Umkehr layers. Vertical lines show the dates of satellite records in COH. The largest impact of the AK is seen between 1997 and 2001 where two curves separate in layers 7, 8 and 9, and also after 2001 in layer 9.

Figure A3 demonstrates time series of monthly mean ground-based records the lower stratosphere at 5 stations. The Umkehr data (blue) are compared with the ozonesonde anomalies either interpolated to the Umkehr layer 3 (green), or ozonesonde profiles matched with Umkehr profiles in time and smoothed using the Umkehr averaging kernels (crimson). All three datasets have been deseasonalized using their respective climatological (using 1998-2008 climatology) average monthly mean ozone. The application of the Averaging Kernels has the effect of smoothing the temporal variability.

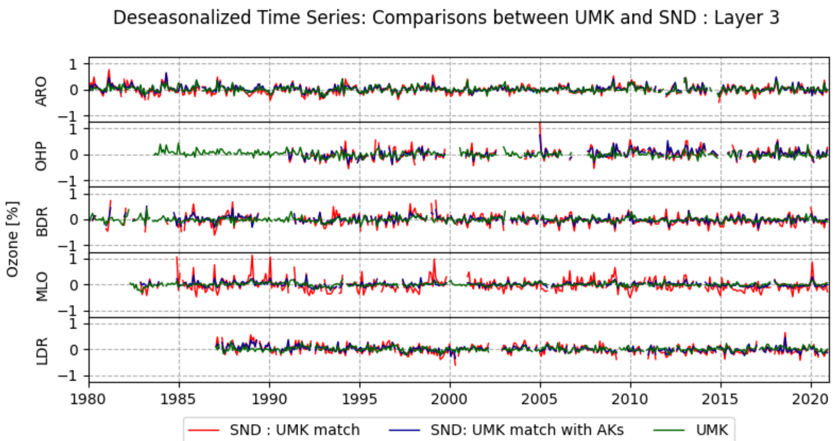


Figure A3: Time series of monthly averaged and de-seasonalized (in %) ozone anomalies of Umkehr (green) and ozonesondes records are compared at 5 ground-based stations. Ozonesonde data are either calculated using only profiles that are interpolated in Umkehr layer 3 (blue) or matched with Umkehr profile in time and smoothed with the Umkehr averaging kernels (crimson).

Appendix B: COH using OMPS v3r2 vs OMPS v4r1

OMPS SNPP v4r1 uses updated SDRs as input which incorporate unified and consistent calibration algorithms removing artificial jumps caused by operational changes, instrument anomalies, or contamination for anomaly views of the environment or spacecraft. Also included are new interpolated band-passes, and updated soft calibration based on the new input SDR's.

Differences between the v3r2 and v4r1 versions of the resulting COH dataset are typically less than 1 percent (Fig. A4 and A5). Small seasonal variation is apparent at all levels. Larger differences are visible in 2020 when the soft calibration for v3r2 is extended beyond its period of relevance. Figure A6 shows the drift between the two versions. Drift between the datasets is less than +- 1% at all levels. This is a reasonable estimate of the resulting expected trend difference in using the newest COH version as compared to the v3r2 results used in Godin-Beekmann (2022).

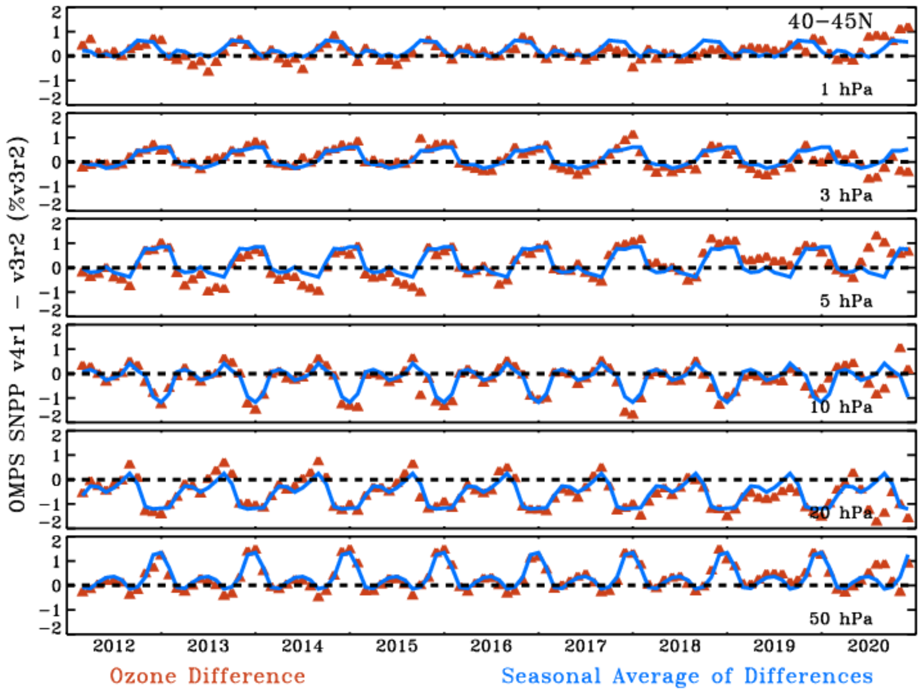


Figure A4: Differences in the COH monthly average zonal product as generated from SNPP v4r1 and v3r2 processing. Also shown is the annual cycle in this difference as depicted by the average over all years for each month. Exhibited at 40-45N is a less than 2% difference with an annual cycle. A somewhat different pattern is seen in 2020 where the soft calibration for v3r2 is extended beyond its period of relevance.

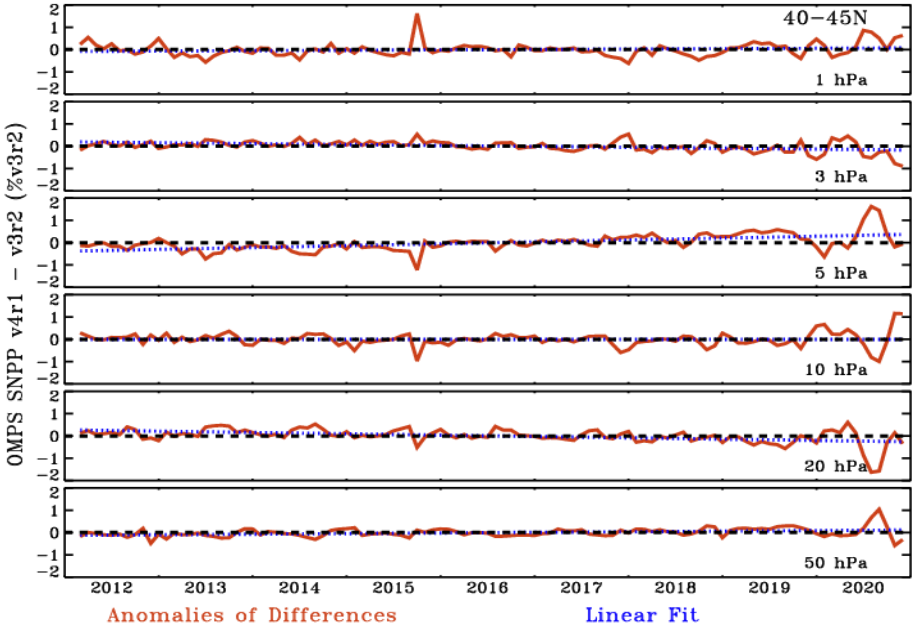


Figure A5: Anomalies of the differences in version (v4r1 vs v3r2) in the COH monthly average zonal product at 40-45N. Anomalies are enhanced in 2020. Also shown as a blue dotted line is a linear least square fit to the anomalies representing the drift between the two versions.

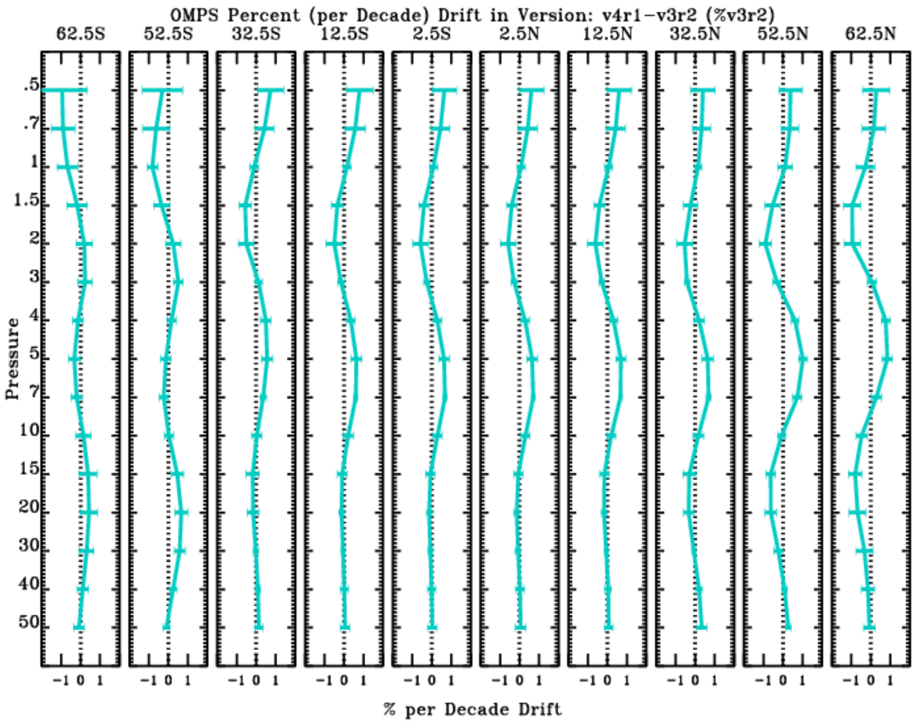


Figure A6 shows the drift between the two versions (v4r1 vs v3r2) as function of pressure level at 10 latitudes.

Appendix C: Impact of using equivalent latitude in generation of the COH product

The COH overpass data used in this paper collects all profiles during the day within a latitude and longitude box of +/- 2 degrees by +/- 20 degrees, then generates a 1/distance averaged value for the station. The box is based on geometric latitude and longitude. With 15 orbits per day, the chosen box size guarantees 2 to 4 possible profiles within the box depending on whether the orbit overpasses or straddles the site as shown in Fig. A7. Also shown is a scenario when the equivalent latitude (EqLat) near the site is particularly non-zonal. In such cases the profiles selected using a geometric coordinate box will select SBUV profiles from an Eq Lat that is different from that of the measurement station.

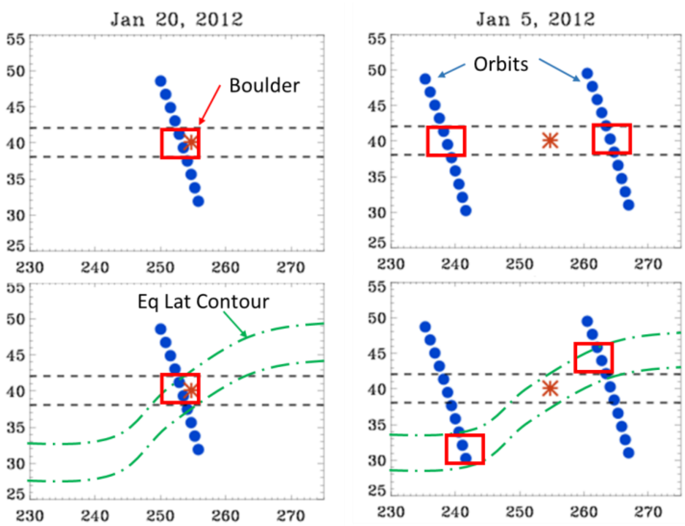


Figure A7: Shows orbits of SNPP and positions of OMPS NP ozone profiles on January 20, 2012 and January 5, 2012. The second row displays a possible EqLat contour overlaid.

It is informative to create an overpass product using boxes based on EqLat and determine the impact on the data. Since EqLat is layer dependent, the included profiles must be selected independently for each layer. Figure A8 shows COH overpass data for Boulder using geometric coordinates, EqLat based coordinates, and the associated Umkehr data. Color coding shows the EqLat at Boulder for each measurement day with dark blue and yellow indicating days with extreme variation from 40N.

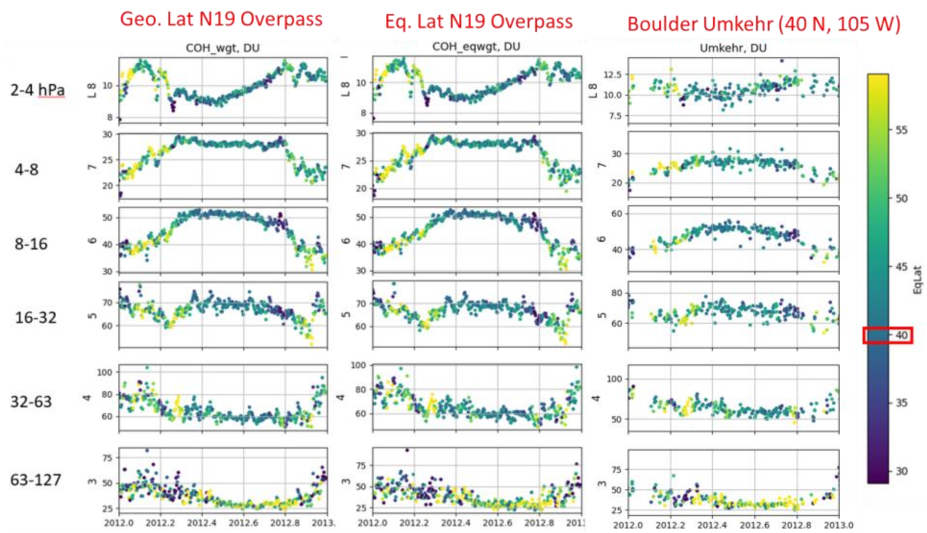


Figure A8: COH overpass data generated with geometric coordinates, EqLat based coordinates, and the associated Umkehr dataset at Boulder for 2012. Data points are color coded for the EqLat at the measurement site. Boulder is at 40 N.

Variation in EqLat is most apparent in Winter months and transitional Fall and Spring, less so in Summer. Yet the value of the COH ozone is not dramatically altered in the time series. Figure A9 shows correlation plots of the COH overpass to Umkehr for the data at layer 7 (4-8 hPa). The pattern of the scatter and the value of the correlation coefficient are not substantially altered for overpass determination using geometric latitude (left) and EqLat (right). Figure A10 shows the vertical distribution of the Correlation coefficient and the RMS Difference for the two COH datasets vs Umkehr. These two metrics are minimally impacted for this sample year in the layers where COH is valid.

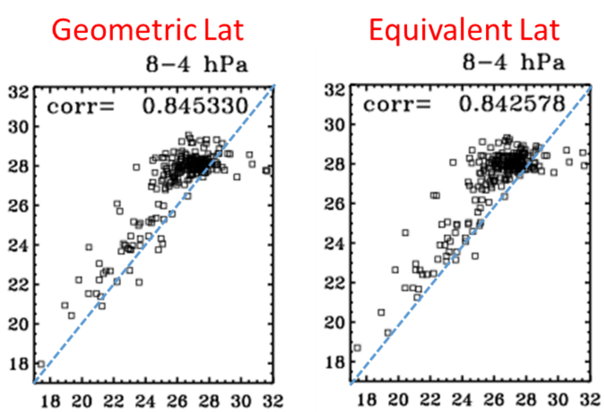


Figure A9: Correlation between Umkehr and COH overpass using Geometric Latitude (left) and EqLat (right) to select included profiles for layer 7 (4-8 hPa).

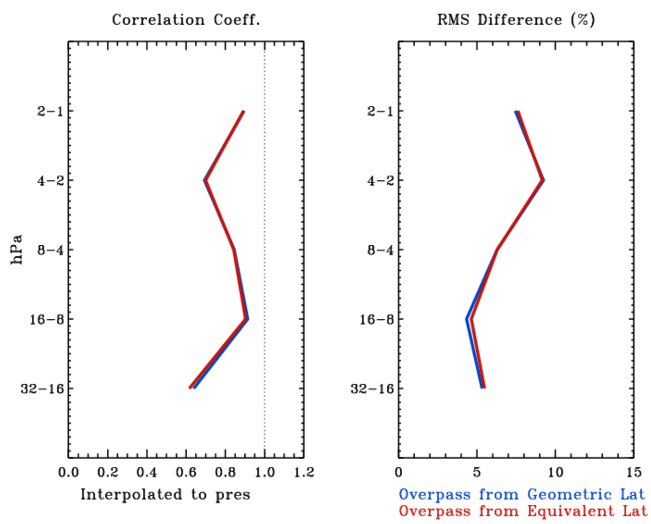


Figure A10: Profiles of Correlation coefficients and RMS differences between COH overpass data at Boulder for 2012 using Geometric Latitude (blue) and EqLat (red) to select data points included in the average.

The use of geometric latitude appears to be sufficient in the choice of included data points in the overpass COH product at the layers used in this paper. Likely this is a ramification of the smooth horizontal resolution of the satellite product.

Appendix D: Temporal Sampling and Impact on Trends

This paper compares trends for three instrument types each with differing measurement frequency. From each set of measurements a monthly average is constructed. See the data files at https://gml.noaa.gov/aftp/ozwv/Publications/2023_Umkehr_Ozone_Trends_Paper/ for the data and the number of data points in each monthly average with the sampling variations. Umkehr measures once or twice per day depending on cloud interference with the measurement. At Arosa/Davos and Lauder, Umkehr measurements are sparser than the other GB stations, often less than 10 per month. At Boulder beginning in 1983 measurements number 20 or more per month. At OHP the Umkehr record begins in 1983 with a strong 20 or more measurements per month. From 1999 to 2016, however, measurements per month are often less than 15 per month. The most Umkehr measurements at MLO are the most abundant, especially after 1985 measuring multiple times in a day, resulting in 50-70 data points contributing to the monthly average. The COH overpass dataset is typically available once per day at each station with occasional misses, contributing usually 27-30 data points per month. Since Umkehr can measure multiple times per day, the COH data matched to Umkehr can contain more profiles in the monthly average than the original full COH data, since the COH overpass data will appear twice in the monthly average, once per each Umkehr measurement.

This occurs often at MLO. Ozonesonde launches are typically one to three times per week depending on the station. At Arosa/Davos, sonde measurements are typically about 15 per month. Sonde measurements at the other stations usually have approximately 5 measurements per month, with some periods of up to 10 per month. As with COH overpass measurements, the sonde dataset matched to Umkehr can have more contributions to the monthly average resulting from dates with more than one Umkehr measurement, resulting in multiple sonde matches.

The trend results in this paper use all available Umkehr data to generate the monthly means. The COH and sonde data are matched to Umkehr to use the Umkehr temporal sampling for COH, and to be able to use the Umkehr averaging kernels for sonde. It is important to determine how the temporal sampling within the monthly mean data may impact trend results. To aid this understanding, we create three subsets of Umkehr data each with different temporal sampling and create the corresponding monthly mean: 1) all observations in Umkehr record; 2) Umkehr matched to the COH dataset; and 3) Umkehr matched to the sonde dataset. In this way we use the same data, but only vary the temporal sampling. Since the COH is measured every day, except in the rare case that the satellite data is missing due to instrument issues, sampling 1 and 2 should provide nearly identical results. We expect a strong change in the monthly mean and resulting trends for Umkehr record when it is matched with infrequent sampling of ozonesonde profiles (especially in Boulder, Hilo and Lauder).

Figure A11 summarizes the results. Each line in Fig. A11 is trend derived from Umkehr data, but with sampling of all data, data matched to COH dates, and data matched to sonde dates. In general, the differences are within the envelope of trend uncertainty (+/- 2 std errors). As expected, the trends and standard errors for all (green) and COH-matched subsampled (orange) Umkehr records are nearly the same. The largest differences in all Umkehr and COH matched Umkehr lines are apparent at OHP. We have determined that this arises from occasional months when there is a short satellite outage coupled with sparse Umkehr observations at the station. However, trends derived from sonde-matched Umkehr data (blue) show deviations from other observations. This is especially clear at Arosa/Davos in the upper stratosphere (~2-3 % above 10 hPa). But since this is above the measurement capability of the ozonesonde, this will not impact the ozonesonde trend results at Arosa/Davos. At Lauder the most significant differences are seen in layer 3 (2.5%),but unfortunately not in the direction to explain sonde differences in the Lauder trend curves as compared to Umkehr. Smaller differences are seen at other layers (very small, less than 1 %, differences in layers 6 and 4). At OHP small differences of less than 1 % are seen between 50 and 10 hPa, well within error estimates.

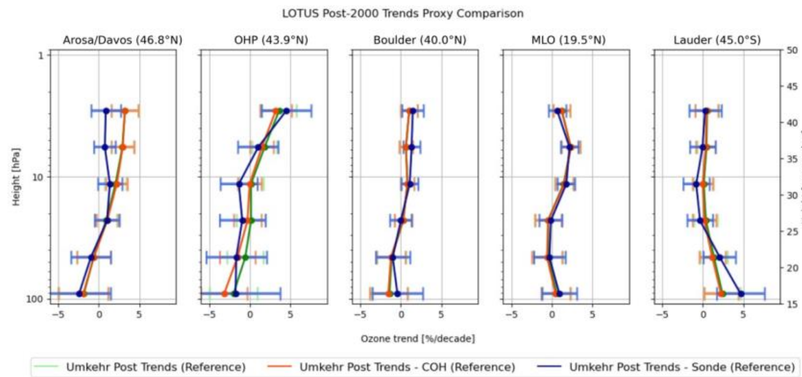


Figure A11: Trend results for the Reference Model using Umkehr data mimicking the temporal sampling of COH and sonde. Green is all available Umkehr data; orange is Umkehr data matched to COH measurements dates; blue is Umkehr data matched to sonde measurement dates.

Figure A12 further explores sampling differences by examination of trends of COH data using the full COH dataset, and data sampled to the Umkehr dates in generation of the monthly mean datasets. As with Fig A11. the trend lines are nearly identical at all stations except OHP. At OHP in the early 2000's there are significantly fewer COH points matched to Umkehr because of the drop in Umkehr measurements. This likely impacts the post-2000 trend estimate. The differences remain below 2%, and are within the error estimate of the trends. In summary, the sampling biases between COH overpass and Umkehr data cannot explain the difference in the derived trends (see Fig. 3, most notable in layers 7 and 8 at Boulder and Lauder).

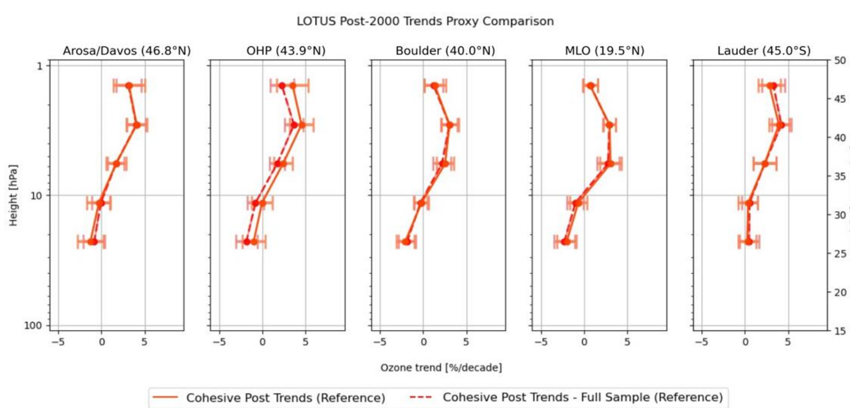


Figure A12: Trend results for the Reference Model exploring variations in sampling of the COH data. Solid orange is COH data matching Umkehr sampling; dotted orange is all available COH data.

Figure A13 explores the impact on trends from sampling differences of the sonde data. Shown are trends with all sonde data, and trends with Umkehr matched data. In this figure only, the sonde data is not AK smoothed since the Umkehr AK are only available on dates when there is an Umkehr measurement. So shown here are trends from sonde data integrated to the Umkehr levels. As with Fig. A11 the only visible impact is seen at OHP and Lauder, though both are within error estimates. At Lauder the trends remain negative for both samplings, but sonde sampled to Umkehr moves closer to the zero line. At OHP the sonde trends are positive, but sonde sampled to Umkehr moves slightly closer to zero. The sampling impact on trends for both OHP and Lauder are likely due to the reduced number of Umkehr data at these sites.

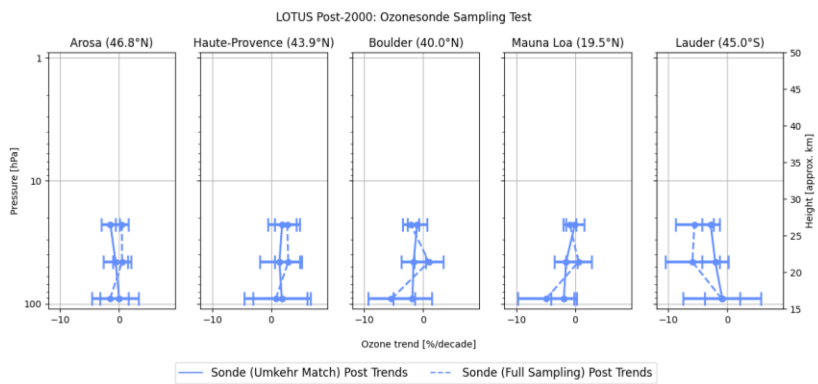
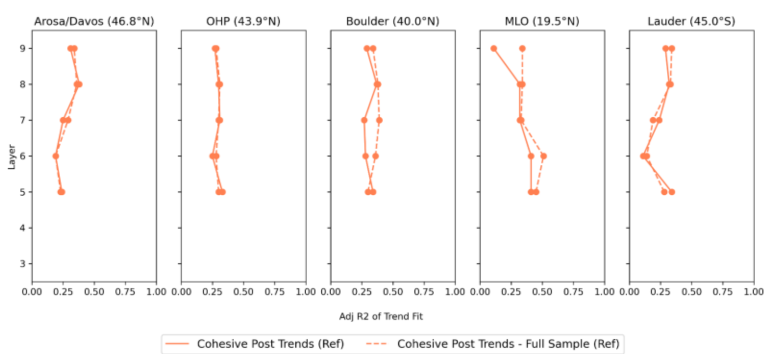
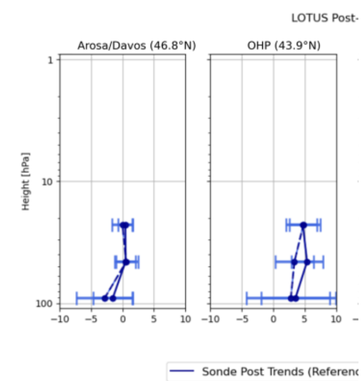


Figure A13: Trend results for Reference model exploring sampling of the sonde data. Solid blue is all sonde data; dashed is Umkehr matched sonde data.

Figure A14 explores the impact of sampling on the adjusted R^2 using the COH overpass data. Shown are the adjusted R^2 for all available COH overpass data, and the same using only COH overpass with matches to the Umkehr data. For Arosa/Davos, OHP and Lauder the differences are small. For Boulder and MLO at some layers (Boulder, layers 6,7; MLO layers 6,9), the impact is more apparent with the Full COH exhibiting higher adjusted R^2 at these stations.





Deleted:

Formatted: English (United States)

Deleted: e

Formatted: Superscript

Formatted: Superscript

Formatted: Superscript

Figure A14: Adjusted R^2 for the Reference Model exploring variations in sampling of the COH data. Solid orange is COH data matching Umkehr sampling; dotted orange is all available COH data.

Appendix E: Decision process for the Full Model

The LOTUS styled Reference Model is developed and optimized for zonal average datasets. The Extended Model tests the addition of single predictors to see if fit statistics can be improved for GB and overpass datasets. For Tropospheric Pressure (TP), improvements are consistent among layers and among instrument types. The addition of EqLat also yields consistent results for instrument types and at most stations, though not Mauna Loa. Addition of other predictors gives mixed results depending on level and station. The potential for improving confidence in trend results exists by combining predictors using different choices depending on layer and station. We choose additional predictor combinations with consideration of three criteria: 1) combined predictors should not have a high correlation with each other (usually .2 or less); 2) predictors should reduce the SE of the trend consistently for all instrument types; 3) addition of the predictor should not greatly reduce the adjusted R^2 of the model fit, but preferentially increase it. As seen in Tables 7e and 7f the NAO and the EHF predictors do not make a significant improvement when added to the Reference Model, so we do not include either in the Full Model.

Mixed Model:

We have noted a high correlation between the TP and EqLat predictors at all levels especially for Boulder, Mauna Loa and Lauder with correlation adjusted R^2 of .4 to .7, and somewhat less correlated at Arosa/Davos and OHP with adjusted R^2 of .2 and .3. Subsequently, we choose to not use these two predictors together (at the same station/layer combination). The addition of TP at all stations for layers 3 and 4 uniformly decreases the standard errors at all stations for both Umkehr and sonde. The addition of EqLat (with the exception of Umkehr at Boulder, level 5) almost uniformly decreases the standard errors at all stations for layers 5 and 6. There is additional reduction in the SE for layers 7 to 9 for all stations except at Mauna Loa. Thus, we choose TP and EqLat as additional predictors at these layers. QBO C and D, have significant impact in decreasing the SE in layers 4 and 5 for both Umkehr and sonde, and layer 3 for sonde with only a small degradation for Umkehr. QBO-CD shows an improvement in layer 8 at OHP, both COH and Umkehr, and Arosa/Davos and Boulder for COH only. We have tested adding both QBO and EqLat for layer 8 at these 3 stations. For Umkehr measurements, there is no improvement beyond EqLat only with QBO-CD also included. For COH there is additional improvement, but not to the extent of QBO-CD alone. Since the improvement is limited to one layer, and for only COH, we choose to only add the additional QBO-CD for the tropical MLO. Table A1 shows the resulting combination of additional predictors for this Mixed Model.

LOTUS Mixed Model					
Layer	**Arosa/Davos**	**OHP**	**Boulder**	**MLO**	**Lauder**
9	EqLat	EqLat	EqLat	Ref	EqLat
8	EqLat	EqLat	EqLat	Ref	EqLat
7	EqLat	EqLat	EqLat	Ref	EqLat

6	EqLat	EqLat	EqLat	EqLat	EqLat
5	EqLat	EqLat	EqLat	EqLat, QBO CD	EqLat
4	TP	TP	TP	TP, QBO CD	TP
3	TP	TP	TP	TP, QBO CD	TP

Table A1: Details of additional predictor combinations for each level and station in the Mixed Model

The resulting change in SE from the Reference Model is shown in Table A2. For most stations/layers this is simply a composite of the values from the single EqLat or TP Extended Model results. There remain a few instrument/station/layers where the SE is slightly increased - Arosa/Davos Umkehr layer 8 and Boulder Umkehr layer 8, but these are negligible. At Boulder Layer 5 Umkehr the increase in SE is somewhat more at 1.85% difference, but this is still small enough to not be of great concern. For Mauna Loa at layers 3,4 and 5 the model is rerun adding two predictors together and the results are new. Indeed, in these cases the SE is improved beyond the single predictor results of either QBO alone, or TP or EqLat alone with the exception of Sonde layer 5 where the change in SE is just slightly degraded from QBO alone (13.42% vs 13.69% reduction in SE).

LOTUS Model Test: Difference [%] in Standard Error: Mixed Model vs Reference Model																	
Pressure	Umkehr	Arosa/Davos			Haute-Provance			Boulder			Mauna Loa			Lauder			
(hPa)	Layer	UMK	COH	SND	UMK	COH	SND	UMK	COH	SND	UMK	COH	SND	UMK	COH	SND	
1-2	9		8.4			2.8			1.9			0.0			2.9		
2-4	8	-0.5	0.7		0.1	1.0		-0.4	1.5		0.0	0.0		1.0	3.1		
4-8	7	3.8	3.0		2.1	1.9		5.4	4.1		0.0	0.0		0.5	1.2		
8-16	6	6.1	8.4		2.5	10.8		2.4	7.8		5.3	0.8		3.4	7.7		
16-32	5	7.9	10.8	5.9	1.9	13.3	8.7	-1.9	0.0	0.9	6.6	11.1	13.4	0.8	3.9		
32-63	4	6.6		6.1	5.9		9.9	3.4		3.0	17.3		9.7	8.0		4.1	
63-127	3	12.8		10.2	12.8		10.7	6.8		2.5	8.0		4.2	9.8		4.4	

Table A2: Change in the SE of the trend using the Mixed Model.

Table A3 shows the adjusted R^2 for the proposed Mixed Models. Similarly to the change in SE (Table A2), the adjusted R^2 is a composite of the individual EqLat or TP results from the extended model with the exception of the results for layers 3, 4, and 5 at Mauna Loa where both predictors are included concurrently. At these layers the adjusted R^2 in some cases matches the higher Adj R^2 values of the two predictors, and in others improves with the combination of QBO and TP or EqLat.

LOTUS Model Proxy Tests: (Adjusted R^2 of Model)																
Height	**Umkehr**	**Arosa/Davos**			**OHP**			**Boulder**			**MLO**			**Lauder**		
(hPa)	**Layer**	**UMK**	**COH**	**SND**	**UMK**	**COH**	**SND**	**UMK**	**COH**	**SND**	**UMK**	**COH**	**SND**	**UMK**	**COH**	**SND**
1-2	**9**		**0.42**			**0.37**			**0.36**			0.11			**0.32**	
2-4	**8**	0.23	**0.39**		0.14	**0.31**		0.17	**0.39**		0.11	**0.32**		0.18	**0.34**	
4-8	**7**	**0.35**	**0.35**		**0.31**	**0.41**		0.27	**0.33**		0.26	**0.32**		0.17	0.27	
8-16	**6**	**0.31**	**0.35**		**0.33**	**0.45**		**0.33**	**0.40**		**0.40**	**0.51**		0.25	0.23	
16-32	**5**	**0.34**	**0.38**	0.26	0.25	**0.51**	0.23	**0.31**	**0.40**	0.19	**0.40**	**0.35**	**0.37**	**0.42**	**0.41**	0.24
32-63	**4**	0.21		0.22	0.29		0.25	0.19		0.13	**0.34**		**0.35**	**0.42**		0.25

Deleted: OHP sonde Layer 5,
Formatted: English (United States)
Deleted: **LOTUS Model Proxy Tests: (% Difference in Std. Error of Model)**
Deleted: associated
Formatted: Superscript
Formatted: Superscript
Formatted: Superscript
Formatted: Superscript
Deleted: ¶
Formatted: English (United States)
Formatted: Superscript
Deleted: 1
Formatted: Font: Not Bold
Deleted: 32

63-127	**3**	0.24		0.23	**0.42**		0.17	0.22		0.13	0.14		0.21	0.25		0.19

Table A3: Adjusted R^2 for the Mixed Model

Augmented Mixed Model

It is hard to ignore the substantial reduction of SE when adding the AO/AAO predictor especially for layers 3,4 and 5 at Mauna Loa, and for layers 3 and 4 at Arosa/Davos. The results for OHP layers 3 and 4 are still compelling, though somewhat less so. So we explore the addition of AO/AAO at these three stations only, for the layers specified. **Table A4** summarizes the predictor choices for this Augmented Mixed Model.

LOTUS Augmented Mixed Model					
	Arosa/Davos	**OHP**	**Boulder**	**MLO**	**Lauder**
Layer					
9	EqLat	EqLat	EqLat	Ref	EqLat
8	EqLat	EqLat	EqLat	Ref	EqLat
7	EqLat	EqLat	EqLat	Ref	EqLat
6	EqLat	EqLat	EqLat	EqLat	EqLat
5	EqLat	EqLat	EqLat	EqLat, QBO, AO/AAO	EqLat
4	TP, AO/AAO	TP, AO/AAO	TP	TP, QBO CD, AO/AAO	TP
3	TP, AO/AAO	TP, AO/AAO	TP	TP, QBO CD, AO/AAO	TP

Table A4: Details of additional predictor choices for each level and station in the Augmented Mixed Model . This differs from Table A1 by adding AO/AAO at some levels for Arosa/Davos, OHP and Mauna Loa.

LOTUS Model Test: Difference [%] in Standard Error: Full Model vs Reference Model																
Pressure	Umkehr	Arosa/Davos			Haute-Provance			Boulder			Mauna Loa			Lauder		
(hPa)	Layer	UMK	COH	SND	UMK	COH	SND	UMK	COH	SND	UMK	COH	SND	UMK	COH	SND
1-2	9		8.4			2.9			1.9			0.0			2.9	
2-4	8	-0.5	0.7		0.1	1.2		-0.4	1.5		0.0	0.0		1.0	3.1	
4-8	7	3.8	3.2		2.1	0.6		5.4	4.1		0.0	0.0		0.5	1.2	
8-16	6	6.1	8.3		2.5	10.9		2.4	7.8		5.3	7.8		3.4	7.7	
16-32	5	7.9	10.6	5.9	1.9	13.4	8.7	-1.9		1.4	13.6	13.0	13.3	0.8	3.9	-1.1
32-63	4	8.7		10.0	6.1		9.4	3.4		2.7	17.3		10.3	8.0		7.4
63-127	3	20.3		18.5	13.5		12.8	6.8		2.2	8.0		5.6	9.8		6.8

Table A5 (the same as Table 10): Change in the SE of the trend using the Augmented Mixed Model.

Table A5 displays the change in the SE from the Reference Model now for the Augmented Mixed Model. Adding AO/AAO at Arosa/Davos (layers 3 and 4) and Mauna Loa (layers 3 to 5) greatly reduces the SE beyond that of the Mixed Model results in **Table A2**. For OHP (layers 3 and 4) the impact is less dramatic for Umkehr. For sonde measurements at layer 4 the AO/AAO addition has no impact beyond the Mixed Model; for layer 3 the addition of AO/AAO results in less reduction of the SE.

Deleted: 26
Formatted: Superscript
Formatted: Font color: Auto
Deleted: LOTUS Model Proxy Tests: (% Difference in Std. Error of Model)

LOTUS Model Proxy Tests: (Adjusted R^2 of the Augmented Mixed Model)																
Height	**Umkehr**	**Arosa/Davos**			**OHP**			**Boulder**			**MLO**			**Lauder**		
(hPa)	**Layer**	**UMK**	**COH**	**SND**	**UMK**	**COH**	**SND**	**UMK**	**COH**	**SND**	**UMK**	**COH**	**SND**	**UMK**	**COH**	**SND**
1-2	**9**		**0.42**			**0.37**			**0.36**			0.11			**0.32**	
2-4	**8**	0.23	**0.39**		0.14	**0.31**		0.17	**0.39**		0.11	**0.32**		0.18	**0.34**	
4-8	**7**	**0.35**	**0.35**		**0.31**	**0.41**		0.27	**0.33**		0.26	**0.32**		0.17	0.27	
8-16	**6**	**0.31**	**0.35**		**0.33**	**0.45**		**0.33**	**0.40**		**0.40**	**0.51**		0.25	0.23	
16-32	**5**	**0.34**	**0.38**	0.26	0.25	**0.51**	**0.23**	**0.31**	**0.40**	0.18	**0.44**	**0.53**	**0.39**	**0.42**	**0.41**	0.29
32-63	**4**	0.23		0.25	0.29		**0.27**	0.19		0.18	**0.42**		**0.38**	**0.42**		**0.31**
63-127	**3**	**0.31**		**0.31**	**0.44**		0.21	0.22		0.11	0.19		0.24	0.25		0.21

Table A6 (The same as Table 12): Adjusted R^2 for the Augmented Mixed Model

Table A6 displays the Adj R^2 for the Augmented Mixed Model. Adding AO/AAO improves the Adj R^2 results for Arosa/Davos and MLO and has little to no impact at OHP. Based on the criteria outlined at the beginning of this appendix, we assign the Augmented Mixed Model as the 'Full Model' in the body of this paper.

Code/Data availability: All dataset used in this study are publicly available at the website https://gml.noaa.gov/aftp/ozwv/Publications/2023_Umkehr_Ozone_Trends_Paper/.

Competing interests: The authors declare that they have no conflict of interest.

Author contributions: IP and JW conceptualized the paper, and IP led the paper preparation. PE, KA, and JW performed the data analysis. KM is responsible for the production of the spatial and temporally matched ground-based and satellite ozone profile data. JW is responsible for producing COH zonally averaged and station overpass ozone profile records. LF is responsible for the retrieval and calibration of the OMPS data. GM, PE, KM and KA are responsible for NOAA Umkehr measurements. EMB is responsible for measurements in Arosa/Davos. RQ is responsible for Umkehr and ozonesonde observations in Lauder, New Zealand. BJ and PC are responsible for ozonesonde observations in Boulder and Hilo. GA is responsible for the ozonesonde observations in OHP. RVM is responsible for HEGIFTOM ozonesonde records and data analyses. RD, SGB, DZ provided context of the LOTUS model use and interpretation of trend analyses. All authors contributed to the writing of the paper.

Acknowledgements:

This study was supported in part by NOAA grant NA19NES4320002 (Cooperative Institute for Satellite Earth System Studies - CISESS) at the University of Maryland/ESSIC and NOAA grant NA22OAR4320151, for the Cooperative Institute for Earth System Research and Data Science (CIESRDS). Additional funding is from NOAA Climate Program Office's Atmospheric Chemistry, Carbon Cycle, and Climate program (AC4), grant numbers NA19OAR4310169 (CU)/ NA19OAR4310171 (UMD). The statements, findings, conclusions, and recommendations are those of the author(s) and do not necessarily reflect the views of NOAA or the U.S. Department of Commerce.

Deleted: **LOTUS Model Proxy Tests: (Adjusted R^2 of Model** ...

Formatted: Superscript

Formatted: Superscript

Formatted: Superscript

The authors would like to thank the NASA/GSFC Atmospheric Chemistry and Dynamics team for the SBUV/2 v8.6 profile data, Eric Beach from the NESDIS/STAR for his help with the S-NPP OMPS data, the NOAA GML observatory team (Boulder, MLO and Fairbanks observatories), LATMOS (OHP), and NIWA (Lauder) for Umkehr and ozonesonde data, Wolfgang Steinbrecht of the DWD for help with interpretation of the Hohenpeißenberg ozonesonde data, the MeteoSwiss and PMOD/WRC teams (Arosa/Davos) for Dobson Umkehr data. Some data are associated with the Network for the Detection of Atmospheric Composition Change (NDACC) and are available through the NDACC website (www.ndacc.org). Additional thanks are due to Susan Strahan for Equivalent Latitude data at each station, to Kai-Lan Chang of CIRES for discussion of statistical interpretation of the thresholds, and to Justin Alsing for development of the LOTUS code. North American Regional Reanalysis (NARR) data provided by the NOAA PSL, Boulder, Colorado, USA, from their website at https://psl.noaa.gov.

References

Anstey, J.A., T.P. Banyard, N. Butchart, L. Coy, P.A. Newman, S. Osprey, and C.J. Wright, Prospect of increased disruption to the QBO in a changing climate, Geophys. Res. Lett., 48 (15), https://doi.org/10.1029/2021gl093058, 2021.

Ancellet, G., Godin-Beekmann, S., Smit, H. G. J., Stauffer, R. M., Van Malderen, R., Bodichon, R., Pazmino, A.: Homogenization of the Observatoire de Haute Provence electrochemical concentration cell (ECC) ozonesonde data record: comparison with lidar and satellite observations, Atmos. Meas. Tech., 15, 3105–3120, https://doi.org/10.5194/amt-15-3105-2022, 2022.

Bai, K., Liu, C., Shi, R. et al. Comparison of Suomi-NPP OMPS total column ozone with Brewer and Dobson spectrophotometers measurements. Front. Earth Sci. 9, 369–380 (2015). https://doi.org/10.1007/s11707-014-0480-5.

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.

Bernet, L., Svendby, T., Hansen, G., Orsolini, Y., Dahlback, A., Goutail, F., Pazmiño, A., Petkov, B., and Kylling, A.: Total ozone trends at three northern high-latitude stations, Atmos. Chem. Phys., 23, 4165–4184, https://doi.org/10.5194/acp-23-4165-2023, 2023.

Bhartia, P. K., Herman, J. R., McPeters, R. D., Torres, O.: Effect of Mount Pinatubo aerosols on total ozone measurements from backscatter ultraviolet (BUV) experiments, J. Geophys. Res., 98, 18,547- 18,554, https://doi.org/10.1029/93JD01739, 1993.

Bhartia, P. K., McPeters, R. D., Flynn, L. E., Taylor, S., Kramarova, N. A., Frith, S., Fisher, B., and DeLand, M.: Solar Backscatter UV (SBUV) total ozone and profile algorithm, Atmos. Meas. Tech., 6, 2533–2548, https://doi.org/10.5194/amt-6-2533-2013, 2013.

Björklund, R., Vigouroux, C., Effertz, P., Garcia, O., 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 tropospheric and stratospheric ozone at Lauder, New Zealand (45S), EGUsphere [preprint], https://doi.org/10.5194/egusphere-2023-2668, 2023.

Boynard, A., Hurtmans, D., Garane, K., Goutail, F., Hadji-Lazaro, J., Koukouli, M. E., Wespes, C., Vigouroux, C., Keppens, A., Pommereau, J.-P., Pazmino, A., Balis, D., Loyola, D., Valks, P., Sussmann, R., Smale, D., Coheur, P.-F., and Clerbaux, C.: Validation of the IASI FORLI/EUMETSAT ozone products using satellite (GOME-2), ground-based (Brewer–Dobson, SAOZ, FTIR) and ozonesonde measurements, Atmos. Meas. Tech., 11, 5125–5152, https://doi.org/10.5194/amt-11-5125-2018, 2018.

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

Chang, K. L., Schultz, M. G., Lan, X., McClure-Begley, A., Petropavlovskikh, I., Xu, X., & Ziemke, J. R., Trend detection of atmospheric time series: Incorporating appropriate uncertainty estimates and handling extreme events, Elem Sci Anth, 9(1), 00035, https://doi.org/10.1525/elementa.2021.00035, 2021.

Chang, K.-L., Schultz, M.G., Koren, G., Selke, N., Guidance note on best statistical practices for TOAR analyses, in TOAR tropospheric ozone assessment report, 2023, available at https://igacproject.org/sites/default/files/2023-04/STAT_recommendations_TOAR_analyses_0.pdf, last access 12/07/2023.

Chang, K.-L., Cooper,O.R., Gaudel, A., Petropavlovskikh, I., Effertz, P., Morris, G., McDonald, B.C., Challenges of detecting free tropospheric ozone trends in a sparsely sampled environment, in Special issue: Tropospheric Ozone Assessment Report Phase II (TOAR-II) Community Special Issue (ACP/AMT/BG/GMD inter-journal SI), 2023 preprint.

Cochrane, D., Orcutt, G. H.: Application of least squares regression to relationships containing auto-correlated error terms,Journal of the American Statistical Association, 44: 245, 32-61,https://doi.org/10.1080/01621459.1949.10483290, 1949 .

DeLuisi, J. J., Longenecker, D. U., Mateer, C. L., Wuebbles, D. J.: An analysis of northern middle-latitude Umkehr measurements corrected for stratospheric aerosols for 1979–1986, J. Geophys. Res., 94, 9837–9846, https://doi.org/10.1029/JD094iD07p09837, 1989.

Diallo, M., Riese, M., Birner, T., Konopka, P., Müller, R., Hegglin, M. I., Santee, M. L., Baldwin, M., Legras, B., and Ploeger, F.: Response of stratospheric water vapor and ozone to the unusual timing of El Niño and the QBO disruption in 2015–2016, Atmos. Chem. Phys., 18, 13055–13073, https://doi.org/10.5194/acp-18-13055-2018, 2018.

Diallo, M. A., Ploeger, F., Hegglin, M. I., Ern, M., Grooß, J.-U., Khaykin, S., and Riese, M.: Stratospheric water vapour and ozone response to the quasi-biennial oscillation disruptions in 2016 and 2020, Atmos. Chem. Phys., 22, 14303–14321, https://doi.org/10.5194/acp-22-14303-2022, 2022.

EEAP (Environmental Effects Assessment Panel): 2022 Quadrennial Assessment - Environmental effects of stratospheric ozone depletion, UV radiation, and interactions with climate change, 2023, https://ozone.unep.org/system/files/documents/EEAP-2022-Assessment-Report-May2023.pdf, last access March 5, 2024.

Evans, R.D., Irina Petropavlovskikh, Audra McClure-Begley, Glen McConville, Dorothy Quincy, and Koji Miyagawa, Technical note: The US Dobson station network data record prior to 2015, re-evaluation of NDACC and WOUDC archived records with WinDobson processing software, Atmospheric Chemistry and Physics, 17, 12051–12070, https://doi.org/10.5194/acp-17-12051-2017, 2017.

Fioletov, V. E., Labow, G., Evans, R., Hare, E. W., Köhler, U., McElroy, C. T., Miyagawa, K., Redondas, A., Savastiouk, V., Shalamyansky, A. M., Staehelin, J., Vanicek, K., and Weber, M.: Performance of the ground-based total ozone network assessed using satellite data, J. Geophys. Res., 113, D14313, https://doi.org/10.1029/2008JD009809, 2008.

Frith, S. M., Kramarova, N. A., Stolarski, R. S., McPeters, R. D., Bhartia, P. K., G. J. Labow: Recent changes in total column ozone based on the SBUV Version 8.6 Merged Ozone Data Set, J. Geophys. Res. Atmos., 119, 9735-9751, https://doi.org/10.1002/2014JD021889, 2014.

Frith, S. M., Kramarova, N. A., Bhartia, P. K., McPeters, R. D., Labow, G. J., J. R. Ziemke, J. R., Haffner, D.: Recent Advances in the SBUV Merged Ozone Dataset (MOD) for LOTUS Phase 2 Analysis of Stratospheric Ozone Trends and Uncertainties, LOTUS Phase 2 Workshop (Virtual), May 28, 2020.

Godin-Beekmann, S., Azouz, N., Sofieva, V., Hubert, D., Petropavlovskikh, I., Effertz, P., Ancellet, G., Degenstein, D., 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.: 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.

Harris, N.R.P., B. Hassler, F. Tummon, G. E. Bodeker, D. Hubert, I. Petropavlovskikh, W. Steinbrecht, J. Anderson, P. K. Bhartia, C. D. Boone, A. Bourassa, S. M. Davis, D. Degenstein, A. Delcloo, S. M. Frith, L. Froidevaux, S. Godin-Beekmann, N. Jones, M. J. Kurylo, E. Kyrölä, M. Laine, S. T. Leblanc, J.-C. Lambert, B. Liley, E. Mahieu, A.

Maycock, M. de Mazière, A. Parrish, R. Querel, K. H. Rosenlof, C. Roth, C. Sioris, J. Staehelin, R. S. Stolarski, R. Stübi, J. Tamminen, C. Vigouroux, K. A. Walker, H. J. Wang, J. Wild, and J. M. Zawodny, Past changes in the vertical distribution of ozone – Part 3: Analysis and interpretation of trends, Atmos. Chem. Phys., 15, 9965-9982, https://doi.org/10.5194/acp-15-9965-2015, 2015.

Hassler, B., I. Petropavlovskikh, J. Staehelin, T. August, P. K. Bhartia, C. Clerbaux, D. Degenstein, M. De Mazière, B. M. Dinelli, A. Dudhia, G. Dufour, S. M. Frith, L. Froidevaux, S. Godin-Beekmann, J. Granville, N. R. P. Harris, K. Hoppel, D. Hubert, Y. Kasai, M. J. Kurylo, E. Kyrölä, J.-C. Lambert, P. F. Levelt, C. T. McElroy, R. D. McPeters, R. Munro, H. Nakajima, A. Parrish, P. Raspollini, E. E. Remsberg, K. H. Rosenlof, A. Rozanov, T. Sano, Y. Sasano, M. Shiotani, H. G. J. Smit, G. Stiller, J. Tamminen, D. W. Tarasick, J. Urban, R. J. van der A, J. P. Veefkind, C. Vigouroux, T. von Clarmann, C. von Savigny, K. A. Walker, M. Weber, J. Wild, and J. M. Zawodny, Past changes in the vertical distribution of ozone – Part 1: Measurement techniques, uncertainties and availability, Atmos. Meas. Tech., 7, 1395-1427, https://doi.org/10.5194/amt-7-1395-2014, 2014.

Hassler, B., Kremser, S., Bodeker, G. E., Lewis, J., Nesbit, K., Davis, S. M., Chipperfield, M. P., Dhomse, S. S., and Dameris, M.: An updated version of a gap-free monthly mean zonal mean ozone database, Earth Syst. Sci. Data, 10, 1473–1490, https://doi.org/10.5194/essd-10-1473-2018, 2018.

Hassler, B., P. Young, P. (Lead Authors), Ball, W. T., Damadeo, R., Keeble,, J., Maillard Barras,E., Sofieva, V., Zeng, G.: Update on Global Ozone: Past, Present , and Future, Chapter 3 in *Scientific Assessment of Ozone Depletion: 2022*, GAW Report No. 278, 509 pp., WMO, Geneva, 2022.

Hubert, D., Lambert, J.-C., Verhoelst, T., Granville, J., Keppens, A., Baray, J.-L., Bourassa, A. E., Cortesi, U., Degenstein, D. A., Froidevaux, L., Godin-Beekmann, S., Hoppel, K. W., Johnson, B. J., Kyrölä, E., Leblanc, T., Lichtenberg, G., Marchand, M., McElroy, C. T., Murtagh, D., Nakane, H., Portafaix, T., Querel, R., Russell III, J. M., Salvador, J., Smit, H. G. J., Stebel, K., Steinbrecht, W., Strawbridge, K. B., Stübi, R., Swart, D. P. J., Taha, G., Tarasick, D. W., Thompson, A. M., Urban, J., van Gijsel, J. A. E., Van Malderen, R., von der Gathen, P., Walker, K. A., Wolfram, E., and Zawodny, J. M.: Ground-based assessment of the bias and long-term stability of 14 limb and occultation ozone profile data records, Atmos. Meas. Tech., 9, 2497–2534, https://doi.org/10.5194/amt-9-2497-2016, 2016.

E. Kalnay, Kanamitsu, M., Kistler, R., Collins, W., Deaven, D., Gandin, L., Iredell, M., Saha, S., White, G., Woollen, J., Zhu, Y., Chelliah, M., Ebisuzaki, W., Higgins, W., Janowiak, J., Mo, K.C., Ropelewski, C., Wang, J., Leetmaa, A., Reynolds, R., Jenne, R., and Joseph, D., The NCEP/NCAR 40-year reanalysis project, Bull. Amer. Meteor. Soc., 77, 437-470, https://doi.org/10.1175/1520-0477(1996)077<0437:TNYRP>2.0.CO;2, 1996.

Knudsen, B. M., and J. U. Grooss, Northern midlatitude stratospheric ozone dilution in spring modeled with simulated mixing. J. Geophys. Res., 105, 6885–6890, https://doi.org/10.1029/1999JD901076, 2000.

Koukouli, M. E., Zara, M., Lerot, C., Fragkos, K., Balis, D., van Roozendael, M., Allart, M. A. F., and van der A, R. J.: The impact of the ozone effective temperature on satellite validation using the Dobson spectrophotometer network, Atmos. Meas. Tech., 9, 2055–2065, https://doi.org/10.5194/amt-9-2055-2016, 2016.

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

Kramarova, N. A., Bhartia, P. K., Frith, S. M., McPeters, R. D., and Stolarski, R. S.: Interpreting SBUV smoothing errors: an example using the quasi-biennial oscillation, Atmos. Meas. Tech., 6, 2089–2099, https://doi.org/10.5194/amt-6-2089-2013, 2013b.

Lary, D.J., Chipperfield, M.P., Pyle, J.A., Norton, W.A. Riishøjgaard, L.P. (1995), Three-dimensional tracer initialization and general diagnostics using equivalent PV latitude–potential-temperature coordinates, Q.J.R. Meteorol. Soc., 121: 187-210, https://doi.org/10.1002/qj.49712152109, 1995.

Lawrence, Z. D., Perlwitz, J., Butler, A. H., Manney, G. L., Newman, P. A., Lee, S. H., & Nash, E. R.,. The remarkably strong Arctic stratospheric polar vortex of winter 2020: Links to record-breaking Arctic Oscillation and ozone loss. *Journal of Geophysical Research: Atmospheres*, 125, e2020JD033271. https://doi.org/10.1029/2020JD033271, 2020.

Lee, H., and Smith, A. K., Simulation of the combined effects of solar cycle, quasi-biennial oscillation, and volcanic forcing on stratospheric ozone changes in recent decades, *J. Geophys. Res.*, 108, D2, 4049, https://doi.org/10.1029/2001JD001503, 2003.

Madronich, S., Lee-Taylor, J. M., Wagner, M., Kyle, J., Hu, Z., & Landolfi, R. (2021). Estimation of skin and ocular damage avoided in the United States through implementation of the Montreal Protocol on Substances that Deplete the Ozone Layer. ACS Earth and Space Chemistry, 5(8), 1876-1888. https://doi.org/10.1021/acsearthspacechem.1c00183.

Maillard Barras, E., Haefele, A., Stübi, R., Jouberton, A., Schill, H., Petropavlovskikh, I., Miyagawa, K., Stanek, M., and Froidevaux, L.: Dynamical linear modeling estimates of long-term ozone trends from homogenized Dobson Umkehr profiles at Arosa/Davos, Switzerland, Atmos. Chem. Phys., 22, 14283–14302, https://doi.org/10.5194/acp-22-14283-2022, 2022.

Manney, G. L., and M. I. Hegglin, 2018: Seasonal and Regional Variations of Long-Term Changes in Upper-Tropospheric Jets from Reanalyses. J. Climate, 31, 423–448, https://doi.org/10.1175/JCLI-D-17-0303.1.

Manney, G. L., and Coauthors, Unprecedented Arctic ozone loss in 2011. Nature, 478, 469–475, https://doi.org/10.1038/nature10556, 2011.

McPeters, R. D., Bhartia, P. K., Haffner, D., Labow, G. J., Flynn, L.: The version 8.6 SBUV ozone data record: An overview, J. Geophys. Res. Atmos., 118, 8032-8039, https://doi.org/10.1002/jgrd.50597, 2013.

Meng, L., Liu, J., Tarasick, D.W.,Randel, W.J., Steiner, A.K., Wilhelmsen, H., Wang, L., Haimberger, L., Continuous rise of the tropopause in the Northern Hemisphere over 1980–2020, Sci. Adv. 7, eabi8065, https://doi.org/10.1126/sciadv.abi8065, 2021.

Millán, L. F., Hoor, P., Hegglin, M. I., Manney, G. L., Boenisch, H., Jeffery, P., Kunkel, D., Petropavlovskikh, I., Ye, H., Leblanc, T., and Walker, K.: Exploring ozone variability in the upper troposphere and lower stratosphere using dynamical coordinates, EGUsphere [preprint], https://doi.org/10.5194/egusphere-2024-144, 2024.

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

Petropavlovskikh, I., Miyagawa, K., McClure-Beegle, A., Johnson, B., Wild, J., Strahan, S., Wargan, K., Querel, R., Flynn, L., Beach, E., Ancellet, G., and Godin-Beekmann, S.: Optimized Umkehr profile algorithm for ozone trend analyses, Atmos. Meas. Tech., 15, 1849–1870, https://doi.org/10.5194/amt-15-1849-2022, 2022.

Rodgers, C. D.: Inverse Methods for Atmospheric Sounding : Theory and Practice., Series on Atmospheric, Oceanic and Planetary Physics, World Scientific Publishing Company, Hackensack, N. J., https://doi.org/10.1142/3171, 2000.

Savin, N. E., & White, K. J. (1978). Testing for Autocorrelation with Missing observations. Econometrica (Pre-1986), 46(1), 59.

Smit, H. G. J. and the ASOPOS panel (Assessment of Standard Operating Procedures for Ozonesondes): Quality assurance and quality control for ozonesonde measurements in GAW, World Meteorological Organization, GAW Report #201, Geneva, Switzerland, available at: https://library.wmo.int/doc_num.php?explnum_id=7167, 2014.

Smit, H. G. J., Thompson, A. M. and the ASOPOS 2.0 panel (Assessment of Standard Operating Procedures for Ozonesondes, v2.0): Ozonesonde measurement principles and best operational practices, World Meteorological Organization, GAW Report #268, Geneva, Switzerland, available at: https://library.wmo.int/doc_num.php?explnum_id=10884, 2021.

SPARC/IO3C/GAW, SPARC/IO3C/GAW Report on Long-term Ozone Trends and Uncertainties in the Stratosphere. I. Petropavlovskikh, S. Godin-Beekmann, D. Hubert, R. Damadeo, B. Hassler, V. Sofieva (Eds.), SPARC Report No. 9, GAW Report No. 241, WCRP-17/2018, https://doi.org/10.17874/f899e57a20b, available at www.sparc-climate.org/publications/sparc-reports/sparc-report-no9, 2019.

Staehelin, J., Pierre Viatte, Rene Stübi, Fiona Tummon, and Thomas Peter, 2018, Stratospheric ozone measurements at Arosa (Switzerland): history and scientific relevance, Atmospheric Chemistry and Physics, 18, 6567–6584, https://doi.org/10.5194/acp-18-6567-2018.

Stauffer, R.M., Thompson, A. M., Kollonige, D. E., Tarasick, D. W., Van Malderen, R., Smit, H. G. J., Vömel, H., Morris, G.A., Johnson, B.J., Cullis, P.D., Stübi, R., Davies. J., Yan, M.M., An Examination of the Recent Stability of Ozonesonde Global Network Data, EArth and Space Science, 9, e2022EA002459. https://doi.org/10.1029/2022EA002459, 2022.

Stauffer, R. M., Thompson, A. M., Kollonige, D. E., Komala, N., Al-Ghazali, H. K., Risdianto, D. Y., Dindang, A., Fairudz bin Jamaluddin, A., Sammathuria, M. K., Zakaria, N. B., Johnson, B. J., and Cullis, P. D.: Dynamical drivers of free-tropospheric ozone increases over equatorial Southeast Asia, EGUsphere [preprint], https://doi.org/10.5194/egusphere-2023-2618, 2023.

Stauffer, R. M., Thompson, A. M., Oman, L. D., & Strahan, S. E., The effects of a 1998 observing system change on MERRA-2-based ozone profile simulations. Journal of Geophysical Research: Atmospheres, 124, 7429–7441, https://doi.org/10.1029/2019JD030257, 2019.

Steinbrecht, W., U. Köhler, H. Claude, M. Weber, J. P. Burrows, and R. J. van der A, Very high ozone columns at northern mid-latitudes in 2010, Geophys. Res. Lett., 38, L06803, https://doi.org/10.1029/2010GL046634, 2011.

Sterling, C. W, Johnson, B. J., Oltmans, S. J., Smit, H. G. J., Jordan, A.F., Cullis, P. D., Hall, E. G., Thompson, A. M., Jacquelyn C. Witte, J. C.: Homogenizing and estimating the uncertainty in NOAA's long-term vertical ozone profile records measured with the electrochemical concentration cell ozonesonde, Atmos. Meas. Tech., 11, 3661–3687, https://doi.org/10.5194/amt-11-3661-2018, 2018.

Szelag, 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, Atmospheric Chemistry and Physics, 20, 7035–7047, https://doi.org/10.5194/acp-20-7035-2020, 2020.

Tarasick, D. W., Davies, J., Smit, H. G. J., Oltmans, S. J.: A re-evaluated Canadian ozonesonde record: measurements of the vertical distribution of ozone over Canada from 1966 to 2013, Atmos. Meas. Tech., 9, 195–214, https://doi.org/10.5194/amt-9-195-2016, 2016.

Thompson, A. M., Witte, J. C., Sterling, C., Jordan, A., Johnson, B. J., Oltmans, S. J., Fujiwara, M., Vömel, H., Allaart, M., Piters, A., Coetzee, G. J. R., Posny, F., Corrales, E., Diaz, J. A., Félix, C., Komala, N., Lai, N., Nguyen, H. T. A., Maata, M., Mani, F., Zainal, Z., Ogino, S., Paredes, F., Penha,T. L. B., da Silva, F. R., Sallons-Mitro, S., Selkirk, H. B., Schmidlin, F. J., Stübi, R., Thiongo, K.: First reprocessing of Southern Hemisphere ADditional OZonesondes (SHADOZ) Ozone Profiles (1998-2016). 2. Comparisons with satellites and ground-based instruments, J. Geophys. Res., 122, https://doi.org/10.1002/2017JD027406, 2017.

Thompson, A. M., H. G. J. Smit, J. C. Witte, R. M. Stauffer, B. J. Johnson, G. Morris, P. von der Gathen, R. Van Malderen, J. Davies, A. Piters, M. Allaart, F. Posny, R. Kivi, P. Cullis, N. Thi Hoang Anh, E. Corrales, T. Machinini, F. R. da Silva, G. Paiman, K. Thiong'o, Z. Zainal, G. B. Brothers, K. R. Wolff, T. Nakano, R. Stübi, G. Romanens, G.J. R. Coetzee, J. A. Diaz, S. Mitro, M. Mohamad and S.-Y. Ogino, Ozonesonde Quality Assurance: The JOSIE–SHADOZ (2017) Experience, Bulletin of the American Meteorological Society, 100, 1, 155-171, 10.1175/BAMS-D-17-0311.1, 2019.

Thompson, A. M., Stauffer, R. M., Wargan, K., Witte, J. C., Kollonige, D. E., & Ziemke, J. R.: Regional and seasonal trends in tropical ozone from SHADOZ profiles: Reference for models and satellite products, J. Geophys. Res.-Atmos., 126, e2021JD034691. https://doi.org/10.1029/2021JD034691, 2021.

Torres, O., Herman, J. R., Bhartia, P. K., and Ahmad, Z.: Properties of Mount-Pinatubo Aerosols as Derived from Nimbus-7 Total Ozone Mapping Spectrometer Measurements, J. Geophys. Res.-Atmos., 100, 14043–14055, https://doi.org/10.1029/95JD01224, 1995.

Tweedy, O. V., Kramarova, N. A., Strahan, S. E., Newman, P. A., Coy, L., Randel, W. J., Park, M., Waugh, D. W., and Frith, S. M.: Response of trace gases to the disrupted 2015–2016 quasi-biennial oscillation, Atmos. Chem. Phys., 17, 6813–6823, https://doi.org/10.5194/acp-17-6813-2017, 2017.

Van Malderen, R., Marc A. F. Allaart, M. A. F., Hugo De Backer, H., Herman G. J. Smit, H. G. J., Dirk De Muer, D. D.: On instrumental errors and related correction strategies of ozonesondes: possible effect on calculated ozone trends for the nearby sites Uccle and De Bilt, Atmos. Meas. Tech., 9, 3793–3816, https://doi.org/10.5194/amt-9-3793-2016, 2016.

Wallace, J.M.,The general circulation of the tropical lower stratosphere. Rev. Geophys. Space Phys.,11, 191–222, https://doi.org/10.1029/RG011I002P00191, 1973.

Wallace, J.M., L. Panetta, and J. Estberg, Representation of the equatorial stratospheric quasi-biennial oscillation in EOF phase space. J. Atmos. Sci., 50, 1751–1762, https://doi.org/10.1175/1520-0469(1993)050<1751:ROTESQ>2.0.CO;2, 1993.

Wargan, K., C. Orbe, S. Pawson, et al. J. R. Ziemke, L. D. Oman, M. A. Olsen, L. Coy, and K. E. Knowland. 2018. Recent decline in extratropical lower stratospheric ozone attributed to circulation changes Geophysical Research Letters, https://doi.org/10.1029/2018gl077406.

Wasserstein, R. L., Schirm A.L., Lazar, N.A., Moving to a World Beyond "$p < 0.05$", The American Statistician, 73:sup1, 1-19, https://doi.org/10.1080/00031305.2019.1583913, 2019.

Weber, M., Arosio, C., Coldewey-Egbers, M., Fioletov, V. E., Frith, S. M., Wild, J. D., Tourpali, K., Burrows, J. P., and Loyola, D.: Global total ozone recovery trends attributed to ozone-depleting substance (ODS) changes derived from five merged ozone datasets, Atmos. Chem. Phys., 22, 6843–6859, https://doi.org/10.5194/acp-22-6843-2022, 2022a.

Weber, M., W. Steinbrecht, C. Arosio, R. van der A, S. M. Frith, J. Anderson, L.Ciasto, M. Coldewey-Egbers, S. Davis, D. Degenstein, V. E. Fioletov, L. Froidevaux, D. Hubert, D. Loyola, C. Roth, A. Rozanov, V. Sofieva, K. Tourpali, R. Wang, and J. D. Wild: Stratospheric Ozone [in "State of the Climate in 2020"], Bull. Amer. Meteor., 103 (8), S90–S92, https://doi.org/10.1175/BAMS-D-22-0092.1, 2022b.

Wild, J.D., S.-K. Yang, and C.S. Long, Ozone Profile Trends: An SBUV/2 Perspective (QOS2016-133), in Proceedings of the Quadrennial Ozone Symposium, Edinburgh, Scotland, https://meetingorganizer.copernicus.org/QOS2016/QOS2016-133.pdf, last access February 6, 2024, 2016.

Witte, J. C., Thompson, A. M., Herman G. J. Smit, H. G. J., Fujiwara, M., Posny, F., Coetzee, G. J. R., Northam, E. T., Johnson, B. J., Sterling, C. W., Mohamad, M., Ogino, S-Y., Jordan, A., da Silva, F. R.: First reprocessing of Southern Hemisphere ADditional OZonesondes (SHADOZ) profile records (1998-2015) 1: Methodology and evaluation, J. Geophys. Res., 122, https://doi.org/10.1002/2016JD026403, 2017.

Witte, J. C., Thompson, A., M., Smit, H. G. J., Vömel, H., Posny, F., Stübi, R.: First reprocessing of Southern Hemisphere Additional Ozonesondes (SHADOZ) Profile Records. 3. Uncertainty in ozone profile and total column, J. Geophys. Res., 123(6), 3243-3268, https://doi.org/10.1002/2017JD027791, 2018.

Witte, J. C., Thompson, A. M., Schmidlin, F. J., Northam, E. T., Wolff, K. R., Brothers, G. B.: The NASA Wallops Flight Facility digital ozonesonde record: Reprocessing, uncertainties, and dual launches. J. Geophys. Res., 124, 3565–3582, https://doi.org/10.1029/2018JD030098, 2019.

WMO (World Meteorological Organization), Scientific Assessment of Ozone Depletion: 2014, Global Ozone Research and Monitoring Project-Report No. 55, 416 pp., WMO: Geneva, Switzerland, available at: https://www.esrl.noaa.gov/csd/assessments/ozone/2014/, 2014.

WMO (World Meteorological Organization), Scientific Assessment of Ozone Depletion: 2018, Global Ozone Research and Monitoring Project–Report No. 58, 588 pp., WMO: Geneva, Switzerland, https://library.wmo.int/idurl/4/56362, 2018.

WMO (World Meteorological Organization), Scientific Assessment of Ozone Depletion: 2022, GAW Report No. 278, 509 pp.; WMO: Geneva, Switzerland, https://ozone.unep.org/science/assessment/sap, 2022.

Wohltmann, I., Rex, M., Brunner, D., and Mäder, J., Integrated equivalent latitude as a proxy for dynamical changes in ozone column, *Geophys. Res. Lett.*, 32, L09811, https://doi.org/10.1029/2005GL022497, 2005.

Wolter, K., and M. S. Timlin, El Niño/Southern Oscillation behaviour since 1871 as diagnosed in an extended multivariate ENSO index (MEI.ext). Intl. J. Climatology, **31**, 14pp., 1074-1087, https://doi.org/10.1002/joc.2336, 2011.

Yoon, S., Kotsakis, A., Alvarez, S. L., Spychala, M. G., Klovenski, E., Walter, P., Morris, G., Corrales, E., Alan, A., Diaz, J. A., and Flynn, J. H., Development and testing of a novel sulfur dioxide sonde, Atmos. Meas. Tech., 15, 4373–4384, https://doi.org/10.5194/amt-15-4373-2022, 2022.

Zhang, J., and Coauthors, Influence of the Arctic Oscillation on the Vertical Distribution of Wintertime Ozone in the Stratosphere and Upper Troposphere over the Northern Hemisphere. J. Climate, 30, 2905–2919, https://doi.org/10.1175/JCLI-D-16-0651.1, 2017.

Zerefos, C., Kapsomenakis, J., Eleftheratos, K., Tourpali, K., Petropavlovskikh, I., Hubert, D., Godin-Beekmann, S., Steinbrecht, W., Frith, S., Sofieva, V., and Hassler, B.: Representativeness of single lidar stations for zonally averaged ozone profiles, their trends and attribution to proxies, Atmos. Chem. Phys., 18, 6427–6440, https://doi.org/10.5194/acp-18-6427-2018, 2018.