https://doi.org/10.5194/egusphere-2023-2668
Preprint. Discussion started: 20 November 2023

Intercomparison of long-term ground-based measurements of tropospheric and stratospheric ozone at Lauder, New Zealand (45S)

Robin Björklund[1], Corinne Vigouroux[1], Peter Effertz[2,3], Omaira E. García[4], Alex Geddes[5], James Hannigan[6], Koji Miyagawa[3], Michael Kotkamp[5], Bavo Langerock[1], Gerald Nedoluha[7], Ivan Ortega[5], Irina Petropavlovskikh[2,3], Deniz Poyraz[8], Richard Querel[5], John Robinson[5], Hisako Shiona[5], Dan Smale[5], Penny Smale[5], Roeland Van Malderen[8], and Martine De Mazière[1]

[1]Department of Atmospheric Composition, Belgian Institute for Space Aeronomy (BIRA-IASB), Brussels, Belgium
[2]Cooperative Institute for Research in Environmental Science, University of Colorado- Boulder, Boulder,CO
[3]NOAA Global Monitoring Lab, Boulder, CO
[4]Izaña Atmospheric Research Center (IARC), State Meteorological Agency of Spain (AEMET), Santa Cruz de Tenerife, Spain
[5]National Institute of Water and Atmospheric Research Ltd (NIWA), New Zealand
[6]Atmospheric Chemistry, Observations & Modeling, National Center for Atmospheric Research, Boulder, CO, USA
[7]Naval Research Laboratory, Washington, DC, USA
[8]Royal Meteorological Institute of Belgium, Ringlaan 3, 1180 Uccle (Brussels), Belgium

Correspondence: R. Björklund (robin.bjorklund@aeronomie.be)

Abstract. Long-term ground-based ozone measurements are crucial to study the recovery of stratospheric ozone as well as the trends of tropospheric ozone. This study is performed in the context of the LOTUS (Long-term Ozone Trends and Uncertainties in the Stratosphere) and TOAR-II (Tropospheric Ozone Assessment Report, phase II) initiatives. Within LOTUS, we want to know why different trends have been observed by different ground-based measurements at Lauder. In TOAR-II, intercomparison studies between the different ground-based data sets are needed to evaluate their quality and relevance for trend studies. To achieve these goals, we perform an intercomparison study of total column ozone and multiple partial ozone columns between the ground-based measurements available at the Lauder station from 2000 to 2022, which are the Fourier transform infrared (FTIR) spectrometer, Dobson Umkehr, ozonesonde, lidar, and the microwave radiometer. Our method applies a comparison between observations by taking into account the different a priori profiles, vertical resolution, and temporal sampling of the techniques when appropriate, making use of the averaging kernels and error covariance matrices. Finally, we compare partial columns, defined to provide independent information: one tropospheric and three stratospheric columns. The intercomparison is analyzed using the median of relative differences (the bias) of FTIR with each of the other measurements, the scaled Median Absolute deviation (MAD_s), and a trend of these differences (measurement drift). For the total column we find a high correlation between FTIR and Umkehr with a Pearson correlation coefficient of 0.98. The total column shows a bias and strong scatter well within the combined systematic and random uncertainties respectively. There is however a drift of 0.6 ± 0.5 %/decade if we consider the full time series. In the troposphere we find a low bias of -1.9% with the ozonesondes, but there is a larger value of bias (10.7%) with Umkehr due to low degrees of freedom for signal here. No drift is found between the three instruments in the troposphere, which is good for trend studies within TOAR-II. In both the lower and upper stratosphere, we get a negative bias for all instruments with respect to FTIR (between -1.2% and -6.8%), but all are within the range of the systematic uncer-

https://doi.org/10.5194/egusphere-2023-2668
Preprint. Discussion started: 20 November 2023

tainties. In the middle stratosphere we seem to find a negative bias of around -5.2 to -6.6%, pointing towards too high values for FTIR in this partial column. We find no significant drift in the stratosphere between ozonesonde and FTIR for all partial columns. We do observe drift between the FTIR and Umkehr partial columns in the lower and upper stratospheres (2.6±1.1 %/decade and -3.2±0.9 %/decade), with lidar in the midle and upper stratosphere (2.1±0.8 %/decade and -3.7±1.2 %/decade), and with MWR in the midle stratosphere (3.1±1.7 %/decade). These drifts point to the fact that the different observed trends in LOTUS are not due to different sampling, vertical sensitivity or time periods and gaps. However, the difference in trends in LOTUS is reduced by applying a new FTIR retrieval strategy, which changes inputs such as the choice of microwindows, spectroscopy from HITRAN2008 to HITRAN2020, and the regularization method.

1 Introduction

The study of ozone plays a crucial role in understanding the effects of climate change, as well as its impact on human health and plant life (see e.g., WMO, 2018; Brasseur and Solomon, 2005). It is common to distinguish between stratospheric ozone and tropospheric ozone. Stratospheric ozone, commonly referred to as the 'ozone layer', serves as our primary defense against harmful UV radiation. The discovery of the hole in the ozone layer, which is most prominent in the over Antarctica in the austral spring but also present in the Arctic (Solomon, 1999; Manney et al., 2011), emphasized the need for monitoring and reducing the depletion of ozone. The Montreal protocol of 1987 was implemented to reduce the emissions of ozone-depleting substances (ODSs), which are known catalysts for the chemical destruction of ozone. Over the past 30+ years stratospheric ozone has been closely monitored through ground-based, in-situ, and satellite observations. As a result of the decreased stratospheric chlorine levels (Jones et al., 2011), the depletion of the ozone hole has been halted and there are indications of its recovery (Solomon et al., 2016). This recovery is particularly evident in the Antarctic, while in regions between 60°S-60°N, the increase in ozone may be offset by a continued decrease in the lower stratosphere (Ball et al., 2018). Similarly, in the Arctic it is currently unclear if there are any positive ozone trends since 2000 due to the larger dynamical variability complicating the observation of this ozone recovery (WMO, 2018).

Ozone is a greenhouse gas that contributes to global warming (Hansen et al., 1997) and poses a significant threat to human health through its effects on the respiratory system (National Research Council, 1991). Unlike stratospheric ozone, tropospheric ozone has a relatively short atmospheric lifetime of hours to weeks (Stevenson et al., 2006). It does not have any direct emission sources; rather, it is a secondary gas formed by the interaction of sunlight with hydrocarbons and nitrogen oxides, which are emitted by various human-made sources such as vehicles, fossil fuel power plants, and industrial activities (Jacob, 2000). The combination of surface-ozone, ozonesondes, and aircraft measurements, with satellite measurements provides a long-term collection of data to study tropospheric ozone trends. These measurements show an increase of ozone since the 1990s in the tropics, but no clear global pattern for tropospheric-ozone trends since 2000 (Gaudel et al., 2018).

This study is done within the context of LOTUS (Long-term Ozone Trends and Uncertainties in the Stratosphere) and the HEGIFTOM focus working group (Harmonization and Evaluation of Ground Based Instruments for Free Tropospheric Ozone Measurements, http://hegiftom.meteo.be) within the TOAR-II (Tropospheric Ozone Assessment Report phase II) initiative.

https://doi.org/10.5194/egusphere-2023-2668
Preprint. Discussion started: 20 November 2023

The current goal of LOTUS is to gain a comprehensive understanding of ozone trends, including their relationship to altitude and latitude, by thoroughly evaluating uncertainties in trend studies and considering the impact of errors related to the sampling and stability of data sets. Within LOTUS, Godin-Beekmann et al. (2022) use a regression model to obtain trends of the stratospheric ozone vertical distribution. They find significant differences in trends between the measurements at Lauder, New Zealand. Comparisons at Lauder have been done done earlier by Bernet et al. (2020), but only using two of the ground-based measurements (stratospheric lidar and microwave radiometer) together with the Aura Microwave Limb Sounding satellite and ERA5 reanalysis data, and thus do not explain the differences for most measurements in Godin-Beekmann et al. (2022). We will look at more of the ground-based measurements which were discrepant in Godin-Beekmann et al. (2022) while additionally extending the study to the troposphere. Concerning tropospheric ozone, the HEGIFTOM working group within TOAR-II focuses in part on the need for a thorough intercomparison of tropospheric ozone measurements, where all biases and drifts of the used instruments are evaluated.

To perform this study, a site with multiple instruments that take near-simultaneous observations is necessary. Here we will work with the ozone-measuring instruments at Lauder, which is a station in the Network for the Detection of Atmospheric Composition Change (NDACC, De Mazière et al., 2018). Our aim is to quantify the biases between the different measurements at this site and potential drifts. This will help us interpret the different trends observed at Lauder (Godin-Beekmann et al., 2022). This study aims to take advantage of the many ground-based ozone-measuring instruments to analyze long-term trends over the past two decades at the Lauder atmospheric research station and to compare their results to track potential biases and drifts. The time series of a number of ozone instruments are used to compare the measurements against each other in the total column, and four different partial columns, one in the troposphere and three in the stratosphere. First, in Section 2, we elaborate on the different ground-based ozone measurements. Next, in Section 3, we explain the intercomparison method where we adjust for the different temporal sampling and vertical resolutions. We select coincident measurements based on the sampling and adjust for differences in vertical resolution by re-gridding and smoothing the profile of the high-resolution instruments. Because we do not have high vertical resolution for all measurements, we divide the vertical ozone profile into 4 partial columns from 0.5 to 11 km, 14 to 21.7 km, 21.7 to 29 km, and 29 to 42 km. In Section 4 we look at measurement total and partial column bias, scatter and drift along with the effect of temporal sampling. We end with a summary and conclusions in Section 5.

2 Ground-based measurements

Five ground-based measurement techniques which have vertical information, available at the Lauder station (45°S, 170°E) are considered in this study: the Fourier transform infrared spectrometer (FTIR), the Dobson Umkehr method, the microwave radiometer (MWR), the stratospheric ozone lidar, and the ozonesonde observations. Some of the specifications for each of these measurements are given in Table 1. Aside from these five measurements, we also consider two total column ozone (TCO) measurement: Dobon TCO (which is separate from the Dobson Umkehr technique) and Bentham ultraviolet double monochromator (UV2, Geddes et al., 2023). In Table 1 we identify a group of vertically resolved measurements, which include

https://doi.org/10.5194/egusphere-2023-2668
Preprint. Discussion started: 20 November 2023

Table 1. Specifications for the ozone measurements available at the Lauder station. These are average, or indicative values for the measurement frequency, altitude boundaries, time period, and degrees of freedom for signal (DOFS, see text for definition) if relevant (otherwise Not Applicable, NA).

	Observing frequency	Vertical extent	Time series	DOFS
FTIR	∼5 per week	Surface to ± 50 km	2001-2022	∼4.5
Dobson	4-5 per week	Surface to ± 50 km	1987-2020	∼3.5
MWR	2-3 per day	± 20 to 50 km	1992–2016	∼7
Lidar	∼1 per week	±10 to 50 km	1994-2011, 2014-2021	NA
Sonde	∼3 per month	Surface to ± 30 km	1986-2022	NA
UV2	∼2 per day	Total column ozone	2012-2022	NA

the ozonesonde and lidar. The second group are instruments with lower vertical resolution (FTIR, MWR, and Umkehr) where profiles are derived from an inversion method by Rodgers (2000). The time series of the total ozone column for all these measurements are shown in Figure 1. Note that only for Dobson TCO and UV2, the actual total ozone column is available. For the other instruments we show the integrated column of the available data which covers only a part of the total ozone profile as we can see in Table 1, which is why the absolute value is shifted downward. The figure shows us the coverage of the observations over the full time span. This displays a more densely sampled time series for the FTIR, Umkehr, MWR, and UV2 measurements as opposed to the less frequent observations of lidar and the ozonesonde. Additionally we see a shorter time span for the MWR data (which stops in October 2016) and for UV2 (which starts in 2012) and a gap of the lidar data between 2012 and 2015. We consider observations made after 2000 to allow for the intercomparison of many ozone measurements and which cover the most important period of stratpospheric-ozone recovery thanks to the reduction of ODSs.

2.1 FTIR measurements

The FTIR instruments record mid-infrared solar transmission spectra at high spectral resolution. The spectra contain the signatures of molecular rotational-vibrational transitions of numerous trace gases (including ozone) in the terrestrial atmosphere as they absorb solar radiation. The spectra are analyzed to measure the concentration of these trace gases in the atmosphere using the pressure and temperature dependence of the absorption line shapes to retrieve vertical profile information from which furthermore a TCO value is calculated. This retrieval strategy is based on the inversion method developed by Rodgers (2000) and the details are described in Vigouroux et al. (2008). An important aspect of the method is the necessity of a priori vertical profiles of both the target gas and interfering species (those species that have absorption lines in the same spectral window as is considered for the analysis of the target gas) as well as the a priori covariance matrices of the target gas (and interfering species). The vertical information is characterized by the averaging kernel $\mathbf{A}$, defined through

$$\mathbf{x} = \mathbf{x}_a + \mathbf{A}\left(\hat{\mathbf{x}} - \mathbf{x}_a\right) + \boldsymbol{\epsilon}_x, \tag{1}$$

https://doi.org/10.5194/egusphere-2023-2668
Preprint. Discussion started: 20 November 2023

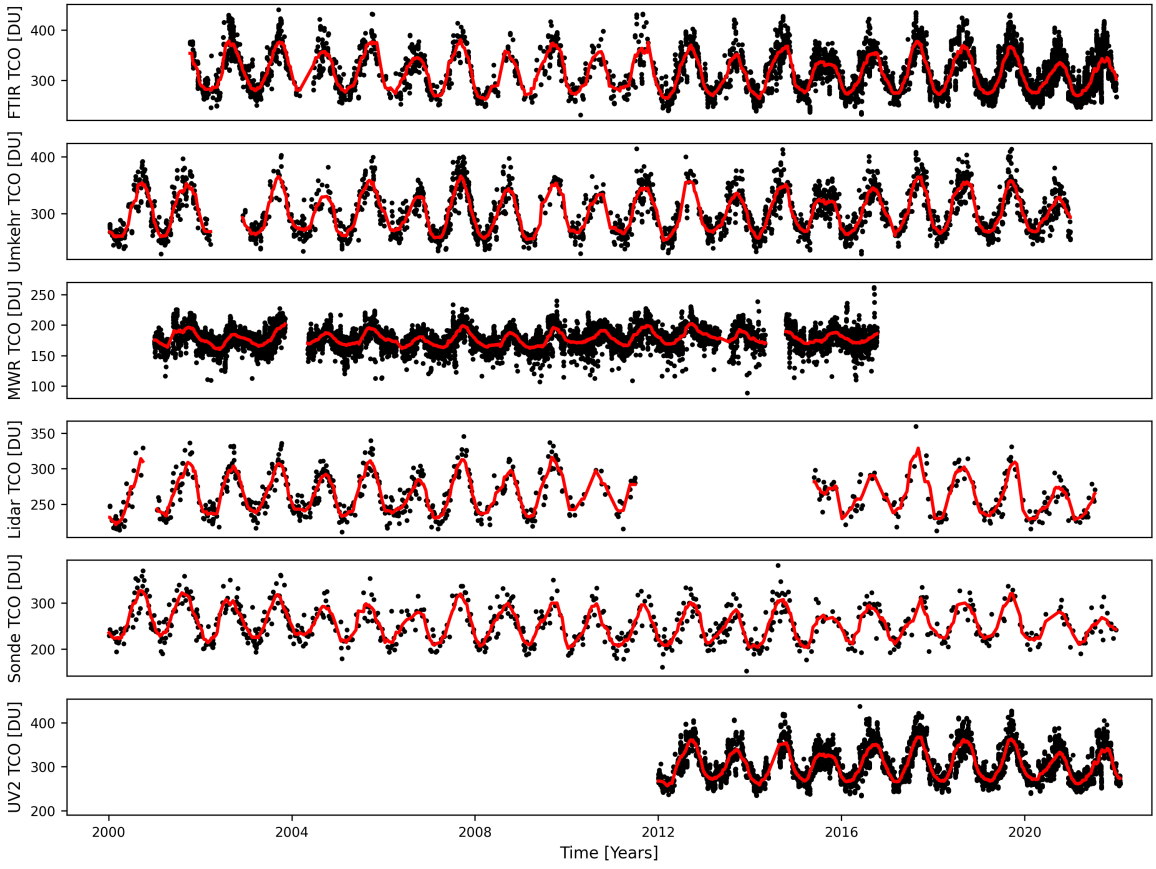

Figure 1. Time series since 2000 of the total ozone column (TCO) in Dobson Units (DU) for FTIR, Umkehr, MWR, lidar, sonde, and UV2 (from top to bottom). The red line shows a running mean of the data with a window of three months. The timeseries for MWR, Lidar, and sonde do not show the full total column of ozone because their limited range of measurement in altitude.

which relates the retrieved state vector $\mathbf{x}$ to the true state $\hat{\mathbf{x}}$ and the a priori state vector $\mathbf{x}_a$, where $\boldsymbol{\epsilon}_x$ contains the measurement and forward model errors. As an example, Figure 2 shows the averaging kernel for one measurement relative to the a priori. Additionally the figure shows the sensitivity of the instrument, which is the sum of the rows of the averaging kernel and indicates the contribution of the measurement to the retrieval compared to the contribution of the a priori. Zero sensitivity would mean that only the a priori contributes to the retrieved profile and the retrieved profile is independent from the real profile. These a priori profiles are additionally shown in Figure 2, which, for FTIR, is the same for every retrieval irrespective of when the spectrum is taken.

The selection of lines from the full solar spectrum (also called microwindows), the a priori information, spectroscopic database, etc., all influence the retrieved ozone profile and thus their treatment requires great care and have been harmonized within the InfraRed Working Group (IRWG, www2.acom.ucar.edu/irwg) of the international collaboration NDACC

https://doi.org/10.5194/egusphere-2023-2668
Preprint. Discussion started: 20 November 2023

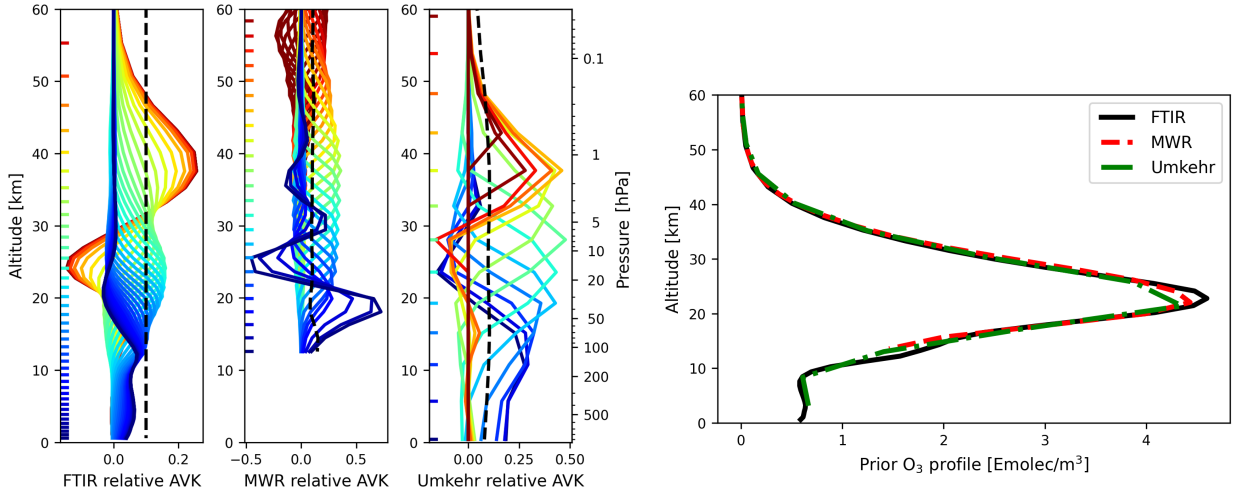

Figure 2. Top figure: Averaging kernel of the three low-resolution instruments. The color corresponds to the altitude marked on the left-hand side of the figures. The black-dashed line shows the sensitivity (divided by 10 to better fit on the scale of the figure). Bottom figure: A priori profile of ozone for FTIR, MWR, and Umkehr. The FTIR retrievals use one fixed prior, while both MWR and Umkehr have priors changing in time. For the latter two, we show a mean profile over all measurements within the considered time frame.

(Vigouroux et al., 2008, 2015). The amount of vertical information that can be obtained is quantified by the degrees of freedom for signal (DOFS) being the trace of the averaging kernel matrix. The total DOFS depend on this retrieval strategy, but of course also on which gas is retrieved and the instrument with which it is observed. For the ozone measurements at Lauder, the total DOFS are typically around 4 to 4.5.

A new retrieval strategy for ozone is used in this work, which is explained in more detail in Section A.

For each of the measurements, we need to consider the instrument uncertainties during the intercomparison study (see Table 2). These are divided into systematic uncertainties and random uncertainties, which capture the accuracy and the precision of the measurement respectively. The source of the random uncertainties are the measurement noise, errors in the forward model parameters, and the smoothing error from the low vertical resolution. This smoothing error, however, is considered separate from the other errors as suggested by the NDACC IRWG. For FTIR, the biggest contribution to the random uncertainty is typically from the temperature profile (Vigouroux et al., 2008). In the case of the systematic uncertainty, the largest contribution actually comes from the spectroscopic database. Specifically the ozone line strength parameter has the biggest uncertainty in the FTIR retrievals. We set it to be 3% (Gordon et al., 2022). As explained in more detail in García et al. (2012), the assumed uncertainty sources are propagated to an uncertainty on the retrieved profile following the technique of Rodgers (2000), which is captured by a full error covariance matrix $\mathbf{S_x}$. Additionally, the smoothing error is calculated from the averaging kernel and the a priori covariance matrix $\mathbf{S_a}$ through

$$U_{\text{smooth}} = (\mathbf{A} - \mathbf{I})\,\mathbf{S_a}\,(\mathbf{A} - \mathbf{I})^T, \tag{2}$$

https://doi.org/10.5194/egusphere-2023-2668
Preprint. Discussion started: 20 November 2023

with $\mathbf{I}$ the unit matrix, and $\mathbf{S_a}$ is obtained from the Whole Atmosphere Community Climate Model (WACCM, Marsh et al., 2013) as an estimation for the error covariance. With this information we can assess which errors dominate at which altitudes, which we have calculated for one year worth of data at Lauder. We see that the smoothing error always has the largest contribution in the troposphere. Concerning the other random errors, the uncertainty on the temperature profile has the second biggest contribution followed by the measurement noise. Above 25 km, the random uncertainty is usually dominated by the temperature profile. As for the contributions to the systematic uncertainty, as mentioned above, the largest contribution comes from the spectroscopic parameters; most notably the line strength parameter (around 3%) followed by the spectroscopic pressure and temperature broadening parameters respectively (around 1-2%). The uncertainty on the temperature profile also plays a role in the contribution to the total systematic uncertainty, starting to become more important above 25 km.

2.2 Microwave radiometer

The microwave ozone radiometer at Lauder measures the spectrum produced by a thermally excited rotational ozone transition at 110.836 GHz. Similar to the FTIR retrieval method described above, the profile retrieval makes use of the change in pressure broadening as a function of altitude. The experimental technique was described in Parrish et al. (1992), and technical details on the instrument used for this work are given in Parrish (1994).

A formal error analysis for the microwave ozone measurements was presented in Connor et al. (1995), and error estimates for Lauder were updated in Tsou et al. (2000). The net precision was determined to be 5-7% between 56 and 1.3 hPa, and the accuracy 7-9%. The vertical resolution of the Lauder measurements is ∼6-8 km (FWHM, full-width at half-maximum) near 10 hPa. Vertical profiles are provided up to 68 km, however retrievals in the mesosphere have somewhat degraded accuracy, precision, and vertical resolution relative to those in the stratosphere. Averaging kernels for the measurements from Lauder are shown in Nedoluha et al. (2015).

2.3 Dobson Umkehr and TCO

The Umkehr method of estimating the vertical profile of ozone using the Dobson ozone spectrophotometer has been performed since the early 1930s (Walshaw, 1989; Gotz et al., 1934) and the Umkehr measurements stand as some of the longest vertical profile records of ozone collected. The Umkehr method is performed using the 'C' wavelength pair of the Dobson ozone spectrophotometer taken during zenith sky measurements at 12 nominal solar zenith angles (SZA) across the range of 60-90°. A prism inside the Dobson breaks down the sunlight spectra and a pair of slits rejects light outside a narrow band of chosen wavelengths. The 'C' wavelength pair consists of a 'short' wavelength (centered on 311.5 nm) and a 'long' wavelength (centered on 332.4 nm). In this range of wavelengths of UV light, absorption by ozone decreases with wavelength creating differential observations between strong and non-absorbing solar spectrum. The log of the ratio of measured intensities is called "N-value".The N-value changes between 90 and 60° SZA as the sun rises or sets and the instrument records the so-called "Umkehr" curve that is sensitive to the vertical ozone distribution above the instrument.

The current NOAA operational Umkehr retrieval algorithm (Petropavlovskikh et al., 2005) uses the Bass and Paur (1985) absorption cross section with its temperature dependence. The profiles are derived using optimal estimation technique (Rodgers,

https://doi.org/10.5194/egusphere-2023-2668
Preprint. Discussion started: 20 November 2023

2000) that includes the a priori ozone profile from ozone climatology (McPeters and Labow, 2012), measurement error and a priori uncertainties which defines the averaging kernels and vertical resolution of retrieved profiles (see Figure 2 for an example of averaging kernels for Umkehr).

The Lauder Umkehr record is homogenized by the method described in Petropavlovskikh et al. (2022) to create a coherent record for the long-term trend analysis of monthly mean anomalies. The method includes a correction to the Umkehr profiles for the stray light that is outside of the nominal band-pass of the Dobson instrument but is not completely filtered out by slits. The record is homogenized using the MERRA-2 GMI (Modern-Era Retrospective Analysis for Research Applications version 2 Global Modeling Initiative) model as a reference to minimize step changes in the record caused by a change in the optical characteristics of the instrument. Individual N-values have a correction factor at each SZA as described in Petropavlovskikh et al. (2022).

A full uncertainty analysis of the Umkehr observations at Lauder is currently not available, but we can estimate the uncertainties from similar measurements at Boulder, USA, using the same technique (see Table 2 for the random and systematic uncertainties used in this study). This provides us with an analysis of the measurement noise as well as the smoothing error in function of altitude. This measurement noise is calculated from simulated MERRA-2 GMI profiles with added random noise based on the error matrix of Umkehr observations. Considering that the measurement noise is only part of the random error in the remote observations, this estimated uncertainty is an underestimation of the true random error. The smoothing error is calculated using Equation (2) using $\mathbf{S_a}$ from the climatology described in Kramarova et al. (2013).

Umkehr total column ozone (from now dubbed Umkehr TCO) is integrated from the vertical profiles supplied by the Umkehr retrieval algorithm. The retrieved integrated Umkehr TCO is constrained by the algorithm to be within the measurement uncertainty of the observed TCO from the Dobson measurements. Additionally, regular TCO measurements are made with the Dobson ozone spectrophotometer (from now dubbed Dobson TCO) using the log of the ratio of solar irradiances made during either direct sun and zenith sky measurements made at regular nominal times (usually one near local solar noon, one in the morning and one in the afternoon). Dobson TCO is derived typically from the A (305.5/325 nm) and D (317.5/339.9 nm) or C and D wavelength pairs with corrections made for Rayleigh scattering and airmass (Evans and Komhyr, 2008). The use of double pairs is designed to minimize the influence from aerosol scattering and clouds. The precision of the Dobson measurement from direct sun observations is around 1% with an accuracy estimated to be around 5% (Basher, 1985).

2.4 Lidar

The Lauder stratospheric ozone lidar is a differential absorption lidar (DIAL) that was installed by RIVM (Rijksinstituut voor Volksgezondheid en Milieu) and began measurements in 1994. It relies on two different beams to extract information about the vertical ozone distribution. It used an XeCl excimer laser for the primary 308 nm beam and a 2-metre Raman cell of Hydrogen gas at 1.3 Bar to produce a 353 nm beam. The 308 nm light is moderately absorbed by ozone while the 353 nm light acts as the reference beam. The time-of-flight from pulse emission through to detection gives the vertical distribution information and the abundance is computed from the ratio of 308 nm and 353 nm backscattered light signals. Over the instrument's multi-decadal lifetime, a number of intercomparisons have been staged with the NDACC travelling standard ozone lidar (e.g.,

https://doi.org/10.5194/egusphere-2023-2668
Preprint. Discussion started: 20 November 2023

Keckhut et al., 2004; Bernet et al., 2020). Currently, measurements are available up to July 2021, when the excimer laser failed. Measurements are expected to resume during the 2023 calendar year. Concerning the lidar uncertainties (see Table 2), there is a systematic uncertainty of approximately 3%, which is mainly associated with uncertainty in the determination of the temperature-dependent absorption cross-section differential. The assumed random uncertainty for lidar in this study, on the other hand, comes from the photon-counting noise.

2.5 Ozonesonde

Ozonesondes are small balloon-borne instruments carried by weather balloons and attached to radiosondes to measure the vertical ozone distribution from the surface to altitudes between 30 and 35 km. The ozonesondes launched in Lauder since 1986 are electrochemical concentration cell (ECC) ozonesondes. In these ozonesondes, a small gas sampling pump forces the ambient air through the cells that are filled with a neutral-buffered potassium iodide sensing solution. The principle of the ozone measurement is then based on the titration of ozone in this solution, so that each O_3 molecule causes (ideally) two electrons to flow in the external circuit. This measured electrical current can then be converted to ozone partial pressure, by knowing the gas volume flow rate of and the temperature in the pump. At Lauder, ozonesondes from the two different ECC ozonesonde manufacturers have been launched, and different sensing solution types have been used as well. As biases exist between those different manufacturer ozonesonde types and between different sensing solution types, the Lauder ozonesonde time series had to be corrected for such biases. Additionally, the location of the pump temperature sensor changed throughout the years, which should also be homogenized. These additional processing steps have been taking into account in the reprocessed, homogenized, Lauder ozonesonde time series here, following the Ozonesonde Data Quality Assessment guidelines of Smit and Oltmans (2012). The quoted overall precision and uncertainty of onzonesonde measurements are 3-5% and 5-10% respectively (see WMO/ GAW, 2021). The systematic uncertainties for the ozonesonde meaurements in this study are claimed to be removed thanks to the homogenization process. The random uncertainty on the measured ozone partial pressured for the ozonesonde are shown in Table 2. A mean value of the uncertainties on the profile is taken to represent the uncertainty on the partial columns, which is likely an overestimation of the actual uncertainty giving values between 4 and 7%.

2.6 UV2

Another TCO measurement available at Lauder is UV2 (Geddes et al., 2023), which is based on a Bentham UV spectrometer. UV2 makes alternating measurement of UV global and direct-sun irradiance, the direct sun spectra are then combined with Dobson-like slit functions and used to calculate ozone using the Dobson method. In Geddes et al. (2023), UV2 ozone was shown to closely agree with Dobson TCO, with a mean bias between them of 2.57 DU and a standard deviation of 1.15 DU. The precision of UV2 is therefore assumed to be consistent with the Dobson TCO at 1%, the accuracy, determined by the mean daily standard deviation is also calculated to be approximately 5%. UV2 observations are available from 2012-2022 and throughout the day. We use a similar comparison setup as for Dobson where all the FTIR total column measurements with a 6 hour difference are averaged and compared to the UV2 observation.

https://doi.org/10.5194/egusphere-2023-2668
Preprint. Discussion started: 20 November 2023

3 Intercomparison setup

3.1 Partial column definition

All measurements do not share the same observation time, vertical extent, or vertical resolution. In order to perform a meaningful intercomparison between the various instruments, a consistent validation setup is necessary. For example, because the DOFS of several instruments are low, we do not have well resolved information of the ozone concentration with altitude. Therefore, the vertical range is subdivided into altitude layers, where we measure 'partial columns' of ozone. Similar to the approach in Vigouroux et al. (2008), these partial columns are defined by looking at the vertical information of the reference FTIR measurements as well as the altitude boundaries every other instrument measures at. Ideally, the partial columns should have approximately 1 DOFS to ensure that the retrieved information comes mainly from the measurement and not from the a priori partial columns, meaning that for a total of ∼4.5 DOFS for the FTIR profile, we can define 4 altitude layers. The first layer is defined from the surface (∼0.5 km) to 11 km, the second from 14 km (the onset of most LIDAR observations) to 22 km, the third from 22 km to 29 km, and the last from 29 km to 42 km. This upper limit is chosen such that the partial column covers all lidar observations and no extrapolation is needed, which would be the case if the upper limit would be chosen higher. This way, all FTIR observations have around 1 DOFS per partial column. From the ozone profiles we then integrate the partial columns, ending up with one tropospheric column and three stratospheric columns (from now on labelled 'lower', 'middle', and 'upper' stratosphere). In the troposphere we can compare FTIR, Umkehr and ozonesonde measurements; in the lower stratosphere we additionally have the lidar measurements; in the middle stratosphere we can compare all measurements; and in the upper stratosphere we compare FTIR, Umkehr, MWR and lidar observations. Because of the lower average DOFS of the Dobson Umkehr measurements, the DOFS of the defined partial layers do not always reach one: we get on average 0.5, 0.6, 0.7, and 1.1 from troposphere up to the upper stratosphere.

Table 2 summarizes this division of partial columns and provides information on the DOFS and the median instrument uncertainties between 2001-2022. These uncertainties are divided into systematic uncertainties and random uncertainties (which includes the random smoothing error), which capture the accuracy and the precision of the measurement respectively. For the remote sounding measurements, the uncertainties within the partial column reported in Table 2 are obtained from the uncertainties contained in the error covariance matrix $\mathbf{S_x}$ of the measurement by

$$U_{\mathrm{PC}} = \mathbf{h^T S_x h}, \tag{3}$$

where $\mathbf{h}$ transforms the volume mixing ratio profiles to the appropriate partial columns, having values of zero at altitudes outside the boundaries of partial column. If a full error covariance matrix is not available, such as with the Umkehr observations, we use the available error estimates which we know in function of altitude. These values are then used to calculate the errors for partial columns by taking a weighted average, similar as described above.

https://doi.org/10.5194/egusphere-2023-2668
Preprint. Discussion started: 20 November 2023

Table 2. Total and partial column information showing the DOFS and mean systematic uncertainties (U_{s}), random uncertainties (U_{r}), random smoothing error (U_{smooth}), and total random error ($U_{\mathrm{r,tot}} = \sqrt{U_{\mathrm{r}}^2 + U_{\mathrm{smooth}}^2}$) of all ground-based measurements. The Umkehr uncertainties are an estimate based on similar measurements in Boulder, USA.

	Troposphere 0.5-11 km					Lower stratosphere 14-22 km				
	DOFS	U_{s} [%]	U_{r} [%]	U_{smooth} [%]	$U_{\mathrm{r,tot}}$ [%]	DOFS	U_{s} [%]	U_{r} [%]	U_{smooth} [%]	$U_{\mathrm{r,tot}}$ [%]
FTIR	1.0	9.1	0.7	6.9	6.9	1.0	7.2	1.1	3.5	3.7
Umkehr	0.5	7.5	1.6	2.7	3.1	0.6	4.0	2.0	1.1	2.3
MWR	–	–	–	–	–	–	–	–	–	–
Lidar	–	–	–	–	–	–	∼3	2.4	–	2.4
Sonde	–	0	7.3	–	7.3	–	0	4.3	–	4.3

	Middle Stratosphere 22-29 km					Upper stratosphere 29-42 km				
	DOFS	U_{s} [%]	U_{r} [%]	U_{smooth} [%]	$U_{\mathrm{r,tot}}$ [%]	DOFS	U_{s} [%]	U_{r} [%]	U_{smooth} [%]	$U_{\mathrm{r,tot}}$ [%]
FTIR	0.9	3.5	2.6	2.7	3.7	1.0	9.6	2.5	2.2	3.3
Umkehr	0.7	3.0	0.4	0.7	0.8	1.1	5.0	0.6	1.2	1.3
MWR	1.1	3.6	2.0	2.5	3.2	2.2	2.9	2.2	1.2	2.5
Lidar	–	∼3	1.8	–	1.8	–	∼3	4.5	–	4.5
Sonde	–	0	4.2	–	4.2	–	–	–	–	–

	Total column				
	DOFS	U_{s} [%]	U_{r} [%]	U_{smooth} [%]	$U_{\mathrm{r,tot}}$ [%]
FTIR	4.3	3.2	1.2	0.12	1.2
Umkehr	3.5	2.0	1.1	0.5	1.2
UV2	-	1.0	5.0	-	5.0

3.2 Re-gridding, smoothing, and prior substitution

In order now to perform a comparison between two measurements we need to take into account the vertical resolution of the instruments are not the same. Furthermore, the remote sounding instruments (FTIR, Umkehr, MWR) have different averaging kernels as well as a priori, which are both important to relate the retrieved state to the true state, as is seen in Equation (1). In this study we always perform the comparison of a measurement with respect to FTIR because here we have measurements in all partial columns and the total column. In the simplest case, we need to account for the difference in vertical resolution of the measurements such as when we compare FTIR with ozonesonde or lidar observations. These typically have high vertical resolution, such as the one ozonesonde measurement shown in Figure 3 in black. In the FTIR retrievals, on the other hand,

https://doi.org/10.5194/egusphere-2023-2668
Preprint. Discussion started: 20 November 2023

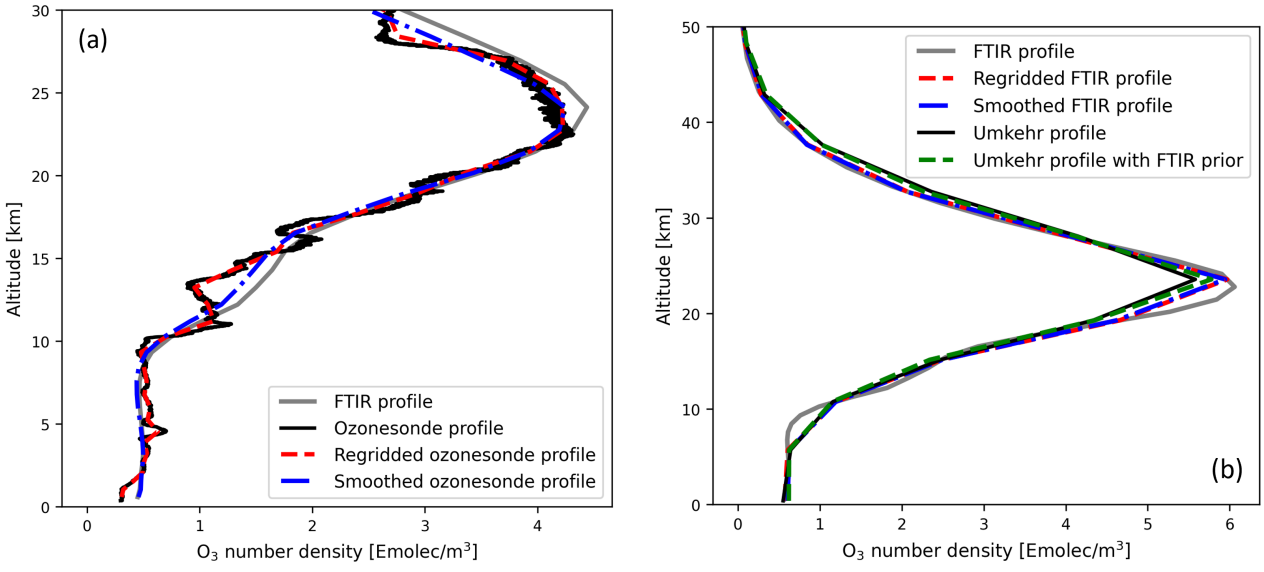

Figure 3. Various stages within the validation method, showing the re-gridding and smoothing of the ozone profile. Figure (a) shows one ozonesonde measurement in comparison with a near-simultaneous FTIR measurement. Figure (b) shows the profiles of Umkehr in comparison with FTIR, where additionally a step of the prior substitution is included.

the altitude grid is discretized into a lot fewer layers (47 for Lauder). To proceed, we will re-grid the measurement of higher resolution to fit the vertical grid of FTIR. For this re-gridding we follow the approach of Langerock et al. (2015) where the 'source' grid of the high-resolution observation is transformed to the FTIR grid. A transformation matrix is constructed which contains the fractions of how the FTIR grid is covered by this source grid, where the interpolation is constructed such that mass is conserved. In Figure 3 (a), an example is shown on how the ozone profile changes due to the re-gridding onto a lower resolution grid. As already explained, the 47 FTIR layers however do not all provide accurate information on the ozone profile because of the limited DOFS of the instrument. In order to simulate this profile retrieval (as would be hypothetically observed by an FTIR measurement) for the higher resolution measurement, the latter is smoothed using the FTIR averaging kernel as described in Rodgers and Connor (2003). By setting the FTIR measurement as 'measurement 1' and the higher resolution measurement as '2', we can smooth the (by now re-gridded) profile $\mathbf{x}_2$ through

$$\mathbf{x}_2^{\text{smooth}} = \mathbf{x}_1^{\text{apr}} + \mathbf{A}_1 \left(\mathbf{x}_2 - \mathbf{x}_1^{\text{apr}}\right), \tag{4}$$

where $\mathbf{x}_2^{\text{smooth}}$ represents the smoothed profile of the higher resolution instrument with the FTIR averaging kernel $\mathbf{A}_1$ using the a priori profile of FTIR $\mathbf{x}_1^{\text{apr}}$. This smoothing step is represented by the blue line in Figure 3 (a) to see the final ozonesonde measurement that is compared to FTIR. The final step is to extract the partial ozone columns by integrating the profiles using the defined partial layers above. This methodology has been applied for the FTIR comparisons with lidar and sondes.

https://doi.org/10.5194/egusphere-2023-2668
Preprint. Discussion started: 20 November 2023

For the remote sounding measurements (MWR and Umkehr) an additional step needs to be considered. These retrievals are typically made with different a priori, which we need to account for in the intercomparison. In the case of the MWR observations, before we smooth the MWR profile using the FTIR averaging kernel, we will transform it, as described in Rodgers and Connor (2003), by substituting the FTIR prior $\mathbf{x}_1^{\mathrm{apr}}$ according to

$$\mathbf{x}_2^{\mathrm{sub}} = \mathbf{x}_2 + (\mathbf{A}_2 - \mathbf{I}) \cdot (\mathbf{x}_2^{\mathrm{apr}} - \mathbf{x}_1^{\mathrm{apr}}), \tag{5}$$

with $\mathbf{I}$ a unit matrix and $\mathbf{x}_2^{\mathrm{apr}}$ the MWR prior. This way the MWR retrieval is adjusted for the different a priori and the profile $\mathbf{x}_2^{\mathrm{sub}}$ is subsequently used in Equation (4), instead of $\mathbf{x}_2$, for the intercomparison after smoothing with the FTIR averaging kernel.

In the case of Umkehr, the same transformation steps apply, however the DOFS are lower than for FTIR so the roles are reversed. This means that the FTIR profiles are first re-gridded to the Umkehr grid (which has 16 vertical layers). Secondly, the Umkehr prior is substituted in the FTIR a priori by Equation (5) where now $\mathbf{x}_2^{\mathrm{apr}}$ is the Umkehr prior. Lastly, before dividing the profile into partial columns, the FTIR profile is smoothed using the Umkehr averaging kernel according to Equation (4), where again here measurements 1 and 2 refer to Umkehr and FTIR respectively. Figure 3 (b) shows the effect each of the transformation steps has on the FTIR and Umkehr profiles. Here we see that the biggest change comes from substitution of the FTIR prior in the Umkehr retrieval, which shifts the Umkehr profile closer to the FTIR profile at the maximum of the ozone number density

3.3 Time coincidence

The observations are not taken exactly simultaneously, so a choice has to be made constructing pairs of observations. The construction is done by considering each separate measurement of Umkehr, MWR, lidar, or ozonesonde, and finding the FTIR measurement which lies within a time window around the observing time of the other instrument, where the window itself depends on the instrument we compare to. The choice of the time windows is elaborated on in Appendix B. By choosing observation pairs as opposed to the full data sets, we automatically select the same time sampling of both instruments when calculating the bias and long-term drift. This leads to a time window of 6 hours for both Umkehr and ozonesondes, 12 hours for lidar, and 3 hours for MWR.

If more than one FTIR measurement falls within the time window of another observation, all these FTIR measurements are averaged. Because we do this, the random uncertainty associated with the measurements will be reduced. If there are N number of measurements within the comparison window, then the random error is reduced by $\sqrt{N}$ for this particular comparison.

4 Results

4.1 Bias and dispersion analysis

Now that we have defined the validation setup to perform the intercomparison, we can analyze the time series of the various ground-based measurements compared to FTIR. The first metric for our validation between the ground-based instruments is

https://doi.org/10.5194/egusphere-2023-2668
Preprint. Discussion started: 20 November 2023

the bias. For these partial columns, we use the observations both of FTIR, as well as the measurements they are compared to. The bias M that we will report here is the median value (where we choose the median over the mean, because of its robust nature with respect to outliers) of the full time series of the relative differences

$$\Delta_{\mathrm{rel}} = \frac{PC_{\mathrm{X}} - PC_{\mathrm{FTIR}}}{PC_{\mathrm{FTIR}}} \cdot 100\% \tag{6}$$

$$M = \mathrm{med}\left(\Delta_{\mathrm{rel}}\right), \tag{7}$$

where we consider the partial column (PC) of FTIR and a second measurement X. Before we use the relative differences to produce the results, we filter for outliers which fall beyond the 3σ deviation. Additionally, to analyze the precision within the intercomparison, we use the MAD (Median Absolute Deviation), which is the median of the absolute deviations from the overall median.

$$\mathrm{MAD}_s = 1.4826 \cdot \mathrm{med}\left(\mathrm{abs}\left(\Delta_{\mathrm{rel}} - M\right)\right). \tag{8}$$

Here is seen that we always use the scaled MAD_s with a constant factor 1.4826. This scaling factor makes the MAD representative as a deviation from the median, similarly as the standard deviation is to the average, in the case of a normal distribution. We are not dealing with perfect Gaussian distributions, but the factor still creates a reasonable value for the scatter. In order to put the bias and MAD_s in perspective, we will compare these to the combined systematic and random uncertainties of the two involved instruments respectively. These combined uncertainties are calculated by

$$\sigma_{\mathrm{comb}} = \sqrt{\sigma_{\mathrm{FTIR}}^2 + \sigma_{\mathrm{X}}^2} \tag{9}$$

where σ can signify either the systematic or random uncertainty. However, if we are dealing with the comparison of two remote sounding instruments (Umkehr and MWR), the simple combination of Equation (9) is not sufficient, because we also need to account for the smoothing error from the retrieval. Because the averaging kernels are not unit matrices, the total combined errors will be correlated. From Rodgers and Connor (2003), we find the covariance of the difference $\mathbf{S}_{\mathrm{comb}}$ to be

$$\mathbf{S}_{\mathrm{comb}} = \left(\mathbf{A}_{\mathrm{FTIR}} - \mathbf{A}_{\mathrm{FTIR}}\mathbf{A}_{\mathrm{X}}\right)\mathbf{S}_{\mathbf{a}}\left(\mathbf{A}_{\mathrm{FTIR}} - \mathbf{A}_{\mathrm{FTIR}}\mathbf{A}_{\mathrm{X}}\right)^T \tag{10}$$

$$+ \mathbf{S}_{\mathrm{FTIR}} + \mathbf{A}_{\mathrm{FTIR}}\mathbf{S}_{\mathrm{X}}{\mathbf{A}_{\mathrm{FTIR}}}^T. \tag{11}$$

This equation includes the separate errors on the covariances $\mathbf{S}_{\mathrm{FTIR}}$ and $\mathbf{S}_{\mathrm{X}}$ as the last two terms and includes the different averaging kernels $\mathbf{A}_{\mathrm{FTIR}}$ and $\mathbf{A}_{\mathrm{X}}$ which account for the smoothing error of both retrievals on the a priori covariance matrix $\mathbf{S}_{\mathbf{a}}$.

For every pair of comparisons we give the numbers of the median and scaled MAD_s of the partial-column relative differences as well as the combined random and systematic uncertainties in Table 3. The results of this table will be discussed in the following sections where we consider the total column and each partial column defined above separately, reporting on these biases and dispersion.

https://doi.org/10.5194/egusphere-2023-2668
Preprint. Discussion started: 20 November 2023

Table 3. Results of the validation for the partial columns which show relative differences $\frac{\mathrm{X-FTIR}}{\mathrm{FTIR}}$ of the measurements (X) with respect to FTIR. The median of the entire time series, or bias M, as well as the scaled MAD_s are shown. For every partial column we also show the combined systematic and random uncertainties of the involved measurements in percent. The last column shows the Pearson correlation coefficient r between the partial column values of FTIR and measurement X.

	Bias M [%]	MAD_s [%]	$\sigma_{\mathrm{sys}}^{\mathrm{comb}}$ [%]	$\sigma_{\mathrm{rand}}^{\mathrm{comb}}$ [%]	r
Sonde					
0.5-11 km	-1.9	6.8	9.1	7.3	0.91
14-22 km	-6.8	4.6	7.2	4.4	0.97
22-29 km	-6.4	4.0	3.5	4.9	0.86
Umkehr					
0.5-11 km	10.7	17.9	11.7	7.6	0.78
14-22 km	-3.2	5.2	8.1	4.3	0.93
22-29 km	-6.6	4.7	4.5	3.8	0.83
29-42 km	-1.9	5.0	9.9	3.6	0.57
TCO	-2.9	1.7	3.2	1.7	0.98
Lidar					
14-22 km	-1.2	8.3	7.8	2.6	0.94
22-29 km	-5.2	4.0	4.6	3.2	0.78
29-42 km	-1.7	4.8	10.1	5.1	0.87
MWR					
22-29 km	-5.7	5.3	6.2	4.9	0.78
29-42 km	-1.4	5.9	10.5	5.2	0.70
Dobson					
TCO	-2.9	2.0	3.2	2.3	0.98
UV2					
TCO	-1.8	2.9	3.4	5.1	0.94

4.1.1 Total column

When applying the intercomparison technique explained in Section 3, we get a full time series of the relative differences between FTIR and all other measurements for the total column and each of the partial columns. The correlation between the FTIR and Dobson total columns shows a very strong Pearson correlation coefficient of 0.98 and similarly high with UV2 (0.94). When we calculate the bias (M) by taking the mean of the relative differences, we get a value of -2.9% and a scaled MAD_s value of 1.7%. These bias and MAD_s values being within the combined systematic and random uncertainties respectively (see Table 3), we can conclude that the total columns from FTIR and Umkehr agree within the uncertainties. Within these

https://doi.org/10.5194/egusphere-2023-2668
Preprint. Discussion started: 20 November 2023

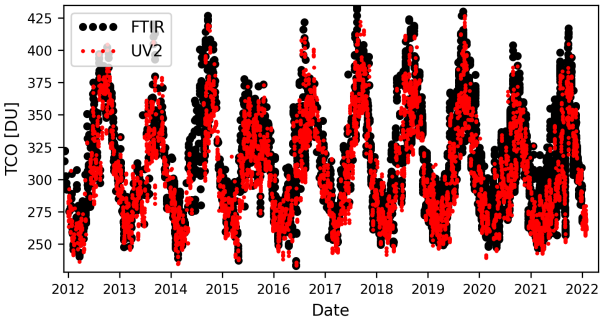

Figure 4. Total column measurements of UV2 and FTIR from 2012 to 2022. Only the direct sun measurements of UV2 are used and show similar behaviour as those from FTIR.

instrument uncertainties we see similar accuracy and precision of both FTIR and Umkehr concerning the total column. The Dobson spectrometer that produces the Umkehr measurements additionally provides TCO during regular solar measurements. Because the instrument is the same as used for the Umkehr data, we do not have a completely independent comparison, however the processing method is different in both cases. For the Dobson TCO we find then very similar values of -2.9% for the bias and 2.0% for MAD_s, which both again are reasonable within the combined instrument uncertainties of Dobson and FTIR.

We can find a similar accuracy and precision of the FTIR measurements when compared to the Brewer measurement (Schneider et al., 2008). Here a larger bias of 4.5% is found between these FTIR and Brewer observations. The improvement to a much smaller bias in this study is largely thanks to the change in spectroscopic input parameters of the FTIR retrieval strategy (see Section A2), where we now use the HITRAN2020 database. Additionally, we can see that FTIR and Dobson has been compared before 2000 by Rinsland et al. (1996) who found a ratio of total columns within 4%.

These results can be supplemented using the UV2 measurements (Geddes et al., 2023). The full UV2 time series and FTIR measurements in the same time window are shown in Figure 4. From the intercomparison of these data sets we find a bias of -1.8% with a MAD_s equal to 2.9%. These values point to a high similarity between the two measurements, both being smaller than the instrument uncertainties (Table 2).

4.1.2 Troposphere

In the tropospheric column between the ground (∼0.5 km) and 11 km, we compare the FTIR observations with both Dobson Umkehr and ozonesonde data. The values of the correlation r for the troposphere is 0.78 and 0.91 for Umkehr and ozonsonde respectively, which are both a strong correlation. In this partial column, we find a small bias of -1.9% of FTIR with ozonesonde and a MAD_s of the relative differences of 6.8%. These values need to be compared to the combined systematic and random uncertainties respectively. In the case of the ozonesonde comparison, it is clear that the found bias is non significant compared to this uncertainty. The MAD_s value also falls within the range of the random uncertainty. For the Umkehr comparison we found a larger bias than for the ozonesonde comparison, but because of the high systematic uncertainties in the, the combined

https://doi.org/10.5194/egusphere-2023-2668
Preprint. Discussion started: 20 November 2023

systematic uncertainty is larger than the bias. Concerning the MAD_s, we again have a larger value of MAD_s = 17.9%. This value is larger than the combined random instrument uncertainties. However, because the Umkehr random uncertainty only accounts for the measurement noise, this combined uncertainty is underestimated. We do notice the effect of the low DOFS value in the troposphere on the high values of both the bias and the MAD_s.

4.1.3 Lower stratosphere

In the lower stratosphere (14-22 km) we can again compare FTIR to Umkehr and sonde data and additionally now the lidar measurements which start above the tropopause. There is a strong to very strong correlation for these partial columns with values of r between 0.93 and 0.97. For all comparisons we find a small negative bias from -1.2% for lidar, -3.2% for Umkehr, and -6.8% for ozonesonde. Similarly as with the results for the troposphere, all these values fall within the range of the combined systematic uncertainties for the respective instruments in the intercomparisons. For the MAD_s of the sonde comparison, we find now value of 4.6%, which is very comparable to the combined random uncertainty. At Izaña for 1999–2010, García et al. (2012) find a bias between FTIR and ozonesondes of -1% with a scatter of around 6% in the range 12–23 km. This bias is smaller than what we find (although both reasonable within the instrument uncertainties), which is partly due to a slightly different altitude range of this partial column, potentially the use of non-homogenized ozonesonde data, and of course because it concerns a different site with a different sampling period. For lidar and Umkehr we get a MAD_s of 8.3% and 5.3%, which are both somewhat larger than the random instrument uncertainties in this partial column, especially for lidar. This could mean that these uncertainties are underestimated, missing or underestimating some contributions to these uncertainties.

4.1.4 Middle stratosphere

In the middle stratosphere (22-29 km) we have all five ground-based measurements available for intercomparison. We again find a strong correlation between the measured partial columns, with r between 0.78 and 0.86. For all measurements we find a negative bias of around -5 to -7% for these measurements with respect to FTIR. The bias found with MWR falls within the combined systematic uncertainties in this column. However, with respect to the ozonesonde, Umkehr, and lidar data, the bias is greater then the combined measurement uncertainties (respectively). All MAD_s values are between 4 and 5.3%, which are slightly larger than the combined random uncertainties. Very similar results are found in García et al. (2012) with a bias of -7% and a scatter of 3% between FTIR and ozonesonde in the 22–29 km altitude range, which they deemed reasonable considering the instrument uncertainties. In our case we do seem to have a remaining bias of FTIR with the other instruments in the middle stratosphere though. When we analyze which elements contribute to the uncertainties in this partial column, the largest contribution with 3.1% is still the spectroscopy. However the uncertainty related to the temperature reaches its highest value here in this partial column, aside from the region above 40 km.

https://doi.org/10.5194/egusphere-2023-2668
Preprint. Discussion started: 20 November 2023

4.1.5 Upper stratosphere

In the upper stratosphere (29-42 km) we can no longer compare to the ozonesonde data, but we still have the other measurements available for intercomparison. The correlation coefficient for the partial columns are moderate (0.57 for Umkehr) to strong (0.70 for MWR and 0.87 for lidar). None of the measurements show a large bias with respect to FTIR where the median value is between -1 and -2% with a MAD_s of 4-6%. So considering respectively the systematic and random instrument uncertainties to compare to these values of bias and MAD_s, the 4 measurements are in agreement. We note though, that even if small, the bias seems to be in FTIR data only, because we find similar values and sign of the bias with all measurements.

4.2 Drift analysis

In our study, we also want to obtain the difference in partial (and total) column trends between the measurements (drift). In the analysis of these relative differences between two measurements within one observation pair, we apply a linear fit to the time series. Detrending of the timeseries is not necessary, because we are dealing with differences of ozone measurements at the same location, so the ozone variabilty should be cancelled out. This way the slope of the linear trend quantifies the drift. We fit the relative differences between monthly means of data in coincidence $Y(t)$

$$Y(t) = A_0 + A_1 t \tag{12}$$

with A_0 the intercept, A_1 gives the drift, and time t is given in fractional years. The monthly means are first calculated from the two measurements separately after which we take the relative difference. Then, the 2-σ trend error obtained from the fit is corrected with the auto correlation of the residuals, according to Santer et al. (2008). This results in a higher uncertainty if any correlation remains in the residuals of the fit. This trend uncertainty has the form

$$U_{\text{drift}} = 2\sigma_{A_1}\sqrt{\frac{N-2}{N_{\text{eff}}-2}}, \tag{13}$$

with σ_{A_1} the standard deviation on the fit parameter A_1, N the degrees of freedom (not to be confused with DOFS from a retrieval), $N_{\text{eff}} = N\frac{1-R}{1+R}$ in which R is the correlation coefficient between the residuals of two consecutive time steps.

To guide the discussion concerning the drift in the total and each of the partial columns, we now summarize the values of these drift for each measurement with respect to FTIR in Table 4. Additionally, Figures 5 and 6 shows the time series of the relative differences within the intercomparison together with the trend analysis. Figure 7 visualizes the obtained drifts in absolute units for the total column as well as all partial columns. These values are obtained by performing the same trend fitting procedure on the absolute differences of the monthly means in Dobson Units (DU). An additional study is done to find the influence of steps in the measurement time series on the calculated drift. These results can be found in Appendix D.

4.2.1 Total column

When we apply the fit from Equation (12) to the relative differences of FTIR and Umkehr we obtain a drift of 0.6 %/decade with an error of $U_{\text{drift}} = 0.4$ %/decade. This means that we find a small drift between FTIR and Umkehr, even when we consider

https://doi.org/10.5194/egusphere-2023-2668
Preprint. Discussion started: 20 November 2023

Table 4. Drifts with respect to FTIR for each of the measurements in %/decade for every partial column where they have data available. The drifts that are significantly different from zero (considering U_{drift}) are represented in boldface.

Drifts [%/decade]	Troposphere 0.5-11 km	Lower stratosphere 14-22 km	Middle stratosphere 22-29 km	Upper stratosphere 29-42 km	Total column
Umkehr	-0.4 ± 6.1	$\mathbf{2.6 \pm 1.1}$	1.5 ± 1.7	$\mathbf{-3.2 \pm 1.2}$	$\mathbf{0.6 \pm 0.4}$
Sonde	-0.1 ± 2.3	-1.4 ± 1.4	0.1 ± 1.4	–	–
Lidar	–	0.4 ± 2.7	$\mathbf{2.0 \pm 0.8}$	$\mathbf{-4.0 \pm 1.2}$	–
MWR	–	–	$\mathbf{3.1 \pm 2.1}$	-1.7 ± 2.1	–
Dobson	–	–	–	–	$\mathbf{0.4 \pm 0.4}$
Dobson	–	–	–	–	$\mathbf{0.3 \pm 1.1}$

the same sampling. Because we have a small, but significant drift we want to compare the FTIR total column to additional TCO measurements at Lauder. The comparison to Dobson TCO is shown in Figure 5. The drift between FTIR and Dobson TCO is 0.4 ± 0.4%/decade. This value is very similar to the one derived from the Umkehr total column, although in this case it is barely non-significant, possibly indicating a small drift the total column measurements from FTIR and Dobson.

Similarly, we can use the relative differences of FTIR with the UV2 data to calculate the drift. For the 2012-2022 period where we have the data, we find a non-significant drift of 0.4±1.8 %/decade. In order to compare this to the Dobson results, we need to recalculate the drift for the same time period. A very similar non-significant drift of 0.3±1.1 %/decade, which points to an agreement between all TCO measurements between 2012 and 2022.

4.2.2 Troposphere

We find non significant drifts of -0.4±6.1 %/decade for Umkehr with FTIR and -0.1±2.3 %/decade for ozonesonde with FTIR, which are shown in Figure 5. Both of these values are non-significant within the trend uncertainty, so we can say with strong confidence that there is no drift of tropospheric ozone between the FTIR, Dobson Umkehr and ozonesonde data sets in the past two decades. These results suggest that at Lauder, we have consistent long-term tropospheric ozone measurements from three independent ground-based measurements, albeit with relatively large uncertainties for the low-resolution instruments.

4.2.3 Lower stratosphere

After performing the intercomparison of the four measurements in this partial column the results show non-significant drifts of -1.0±1.4 %/decade for FTIR with the sonde data and 0.4±2.7 %/decade for FTIR with the lidar data. We do, however, find a positive drift of Umkehr with respect to FTIR of 2.6±1.1 %/decade. In Section 4.3 we will compare how these drifts align with the observed stratospheric trends in Godin-Beekmann et al. (2022).

https://doi.org/10.5194/egusphere-2023-2668
Preprint. Discussion started: 20 November 2023

4.2.4 Middle stratosphere

The resulting drifts between FTIR and the other measurements are 1.5±1.7 %/decade with Umkehr and 0.1±1.4 %/decade with ozonesonde, both non-significant, and two positive drifts of 2.0±0.8 %/decade with lidar and 3.1±2.1 %/decade with MWR. The study of Bernet et al. (2020) also considers both lidar and MWR as two ground-based measurements in the study of stratospheric ozone trends at Lauder. These agree to within 2 to 3 % with each other, which corresponds to similar numbers in our intercomparison. We do not have a full time series available to have the same period of analysis as the Umkehr and ozonesonde comparisons As mentioned above, the lidar measurements are missing 3 years of data from 2012-2015. This incomplete sampling may influence the real lidar trend and potentially the drift with FTIR. Additionally, the MWR data only extends to 2016 at Lauder. This shorter time series also leads to different observed trends, on which we elaborate in Appendix A2.

4.2.5 Upper stratosphere

We find negative drifts for Umkehr -3.2±1.2 %/decade with FTIR and for lidar -4.0±1.2 %/decade with FTIR, and we find a non-significant drift of MWR -1.7±2.1 %/decade with FTIR. From Figure 5 it does seem that the relative differences of lidar and FTIR decrease towards the later part of the time series only. However, as mentioned for the drift in the middle stratosphere, the incomplete time series of the lidar data could lie at the origin of the magnitude of this drift. While in this partial column, the conclusions concerning the drift for lidar and MWR with FTIR differ from one another in magnitude, the values are still in agreement with Bernet et al. (2020), where both measurements have small positive trends in the upper stratosphere around 2 to 3 %/decade.

4.2.6 Drift discussion

The non-zero drifts point to differences between two measurements, where one or both of the instruments could be drifting. The origin of these drifts should be found then in the individual time series of the measurements, tracking all instrument and data processing changes, which is beyond the scope of this paper. Notably though, we see a similar negative drift in the same direction for all instruments in the upper stratosphere indicating a drift in FTIR (as reference measurement for all) in this column. In a first investigation of the FTIR measurements, we analyze the behaviour of the DOFS. We do not, however, find any significant trend in the DOFS time series to help explain the drift. Additionally the signal to noise ratio of the retrievals stays constant over the full time period. Another influence on the drift could be a discontinuous step in the times series caused by instrumental factors, which we discuss in detail in Appendix D. Here we find that a step in May 2018 is present, but taking this into account does not remove the drift. Such a step as found in FTIR shows the need for a detailed instrument log of all instruments to perform similar change-point analyses. While the step in the FTIR data does reduce the drift slightly, it is still present and the reason of this drift and those in the other partial columns still requires some more in depth analysis.

https://doi.org/10.5194/egusphere-2023-2668
Preprint. Discussion started: 20 November 2023

Table 5. Trends for each of the observation measurements in %/decade for every partial column where they have data available. For FTIR, we show the trends derived with both the current and new retrieval strategies and also selecting only the 2000-2016 time period to calculate the trend. The FTIR trends from Godin-Beekmann et al. (2022) (GB22) are additionally shown with their relevant altitude ranges.

Trends [%/decade]	Troposphere 0.5-11 km	Lower stratosphere 14-22 km	Middle stratosphere 22-29 km	Upper stratosphere 29-42 km
Current FTIR	3.9 ± 1.8	-3.6 ± 2.1	-3.2 ± 1.4	4.2 ± 1.1
Current FTIR (2000-2016)			-1.9 ± 2.0	4.1 ± 1.6
New FTIR	1.1 ± 1.7	-3.5 ± 2.0	-2.8 ± 1.0	3.1 ± 1.1
GB22 FTIR	–	-4.5 ± 2.7	-1.7 ± 1.2	5.0 ± 1.1
	–	(12 – 20 km)	(20 – 29 km)	(29 – 49 km)
Umkehr	1.9 ± 4.0	1.7 ± 1.7	-0.4 ± 1.6	0.0 ± 0.9
Sonde	1.5 ± 2.2	-5.1 ± 2.1	-2.6 ± 1.2	–
Lidar	–	-3.0 ± 2.7	1.1 ± 1.3	1.8 ± 2.0
MWR	–	–	2.3 ± 1.6	3.2 ± 1.5

4.3 Comparison to other trend-analysis studies

In order to put the drifts we found in Section 4.2 in perspective, we compare to the trend estimates of Godin-Beekmann et al. (2022). One thing to note is that the FTIR data used in their study uses the current FTIR retrieval strategy, so in order to make a sensible comparison, we should also use the current FTIR results. Furthermore, the FTIR trends are given for partial columns and not for the profile such as is the case for the other measurements, where also the altitude ranges have a different definition then ours. The comparison allows us to, on the one hand, see the differences between using trends in the profile or using trends in the partial columns; and on the other hand we can find the effect of sampling on the trends. The latter matters because Godin-Beekmann et al. (2022) use all measurements to calculate monthly means before deriving trends while here we make a selection of time coincidences in the intercomparison. Table 5 shows the trends calculated for our partial columns using a similar LOTUS regression model as in Godin-Beekmann et al. (2022). Figure 8 visualizes this same information in a comparable way to how the results are presented in (Godin-Beekmann et al., 2022). No drift is found for ozonesonde with FTIR in any of the partial columns, which matches with the similarity in trends from both techniques. However for other techniques, such as in the lower stratosphere we see a positive drift of Umkehr with FTIR. The trends from Godin-Beekmann et al. (2022) here give us approximately -4.5 %/decade for FTIR and between 2.9 and 0.9 %/decade for Umkehr from 16 to 21 km, which are both very similar to the trends in Table 5. Because the trends of FTIR and Umkehr are different and even of opposite sign, this gives confirmation of the drift we find in the lower stratosphere of 2.6±1.1 %/decade. This agreement allows us to say that the trends found in Godin-Beekmann et al. (2022) (which match our trends) are not a matter of sampling or a difference in analyzing trends in partial columns as compared to the profile. In the middle stratosphere we find a positive

https://doi.org/10.5194/egusphere-2023-2668
Preprint. Discussion started: 20 November 2023

significant drift with respect to both lidar and MWR of 3.1±1.7 %/decade and 2.1±0.8 %/decade respectively. In this region, Godin-Beekmann et al. (2022) find trends of the order of -1.7% for FTIR; between -0.4 and 3.3 %/decade for lidar at altitudes from 22.5 km to 28,5 km; and between 2.4 and 1.9 %/decade from 22.5 km to 27 km for MWR which are again similar to the trends we find. Only our trend for FTIR is more negative, but still reasonable considering both trend uncertainties as well as a slightly different altitude range used to derive this trend (20-29 km). If we furthermore mimic the same sampling as the MWR dataset (ending in 2016), the trend seems to better match with the drift. The fact that we find positive drift values of lidar and MWR with respect to FTIR in the middle stratosphere is not surprising considering their opposite sign in trends and thus correcting for the sampling does not seem to change the instrument trends by much. In the upper stratosphere we discuss both Umkehr and lidar drifts. This time the drifts we find are negative (-3.2±0.9 %/decade for Umkehr and -3.7±1.2 %/decade for lidar), because FTIR has a more positive trend in this partial column. These drifts are compared to the Godin-Beekmann et al. (2022) trends of around 5.0 %/decade for FTIR; between -0.8 and 0.2 %/decade from 31 to 37 km for Umkehr; and between 2.8 and -2.1 %/decade from 30 to 42 km for lidar with again similar values in Table 5. The only difference here is that our trend for the lidar observations is bigger, while Godin-Beekmann et al. (2022) show smaller and even some negative trends within this altitude range. This does point to a small difference of profile versus partial columns trends of lidar at this altitude.

5 Conclusions

Long-term measurements of ozone are important to study the recovery of stratospheric ozone as well as the trends of tropospheric ozone. Because many sites use different measurements to monitor these ozone trends, there is a need to validate ozone measurements against each other. In this study we take advantage of the multitude of measurements available at the Lauder station (FTIR, Dobson Umkehr, ozonesonde, lidar, MWR) to perform an intercomparison between these measurements. The study, performed in the context of the LOTUS and TOAR-II initiatives, aims to show the biases and drifts, which require more attention following the results of the trend analysis in Godin-Beekmann et al. (2022).

The method we use applies a comparison between observations (following Rodgers and Connor (2003) for the intercomparison of measurements with different vertical resolution) by manipulating the profiles through prior-substitution, re-gridding, smoothing and finally division into partial columns to perform the comparison. These steps are necessary because of the differences in profile retrieval and vertical resolution of the observations. Additionally, we take care to select pairs of comparisons within a specific time window where multiple FTIR measurements (if present) are averaged before performing the comparison.

For each of the instruments we find a strong correlation with the FTIR partial columns. For the total column, we even find a correlation of 0.98. This shows that, even though a bias or drift might be present, the agreement of the long-term ozone measurements between FTIR and the other measurements is strong, capturing the same variability in all partial layers.

The metrics we use to analyze the intercomparison are the bias and (scaled) MAD_s to track the accuracy and precision of the observations. These are compared respectively to the combined systematic and random instrument uncertainties. We find good agreement of FTIR with Umkehr on the total column concerning the bias (-2.9%) and MAD_s (2.0%) values, similar to Brewer-FTIR comparisons by Schneider et al. (2008). In the troposphere we find a low bias of -1.9% with the ozonesondes,

https://doi.org/10.5194/egusphere-2023-2668
Preprint. Discussion started: 20 November 2023

but there is a larger value of bias (10.7%) and MAD_s (17.9%) with Umkehr due to the low DOFS we have in this column for Umkehr. In both the lower and upper stratosphere we get a small negative bias between -1.2 and -6.8% for all instruments with respect to FTIR, but all fall within the range of the systematic uncertainties. This is not the case, however, for all values of MAD_s (between 4.6 and 8.3 %) which is potentially due to an underestimation of the random instrument uncertainties. In the middle stratosphere we seem to find a negative bias between -5.2 and -6.6%, pointing towards too high values for FTIR in this partial column, possibly related to temperature profile or the treatment of the instrument line shape (ILS). When we look at the uncertainties in this partial column, we notice a higher error from the temperature profile compared to other partial columns. When we compare the results to the intercomparison of FTIR with ozonesonde at Izaña (García et al., 2012), who find similar biases, they are all reasonable within the instrument uncertainties.

We calculated measurement drift by performing a linear fit to the relative differences. This results in a small, but significant drift in the total column of 0.6±0.5 %/decade between FTIR and Umkehr. Additionally some drift is present between these measurements in the lower and upper stratospheres with values of 2.6±1.1 %/decade and -3.2±0.9 %/decade, with lidar in the middle and upper stratosphere of 2.1±0.8 %/decade and -3.7±1.2 %/decade, and with MWR in the middle stratosphere of 3.1±1.7 %/decade. The drifts of the measurements in the same direction could point to different behaviour of FTIR. In comparisons of FTIR and ozonesonde we find no significant drift. This means that there is strong agreement between ozonesonde and FTIR over all partial columns.

When comparing the stratospheric trends of each of the partial column for every measurement, we see that most of the trends are in agreement with those found in LOTUS (Godin-Beekmann et al., 2022), showing that the approach of partial columns in this study does not change the result much from considering the profile itself. Only in the upper stratosphere the trend changes rapidly with altitude, we find a small difference due to this effect, which in fact reduces the discrepancy between the lidar and FTIR trends at this altitude. These trends are, in turn, mostly in agreement with the drift. The reason for some discrepancy might be due to the fact that we account for different sampling of the data by constructing the comparison pairs. One example here is the missing 3-year gap in the lidar data at 2012-2015 or the shorter time series of MWR which stops in 2016.

Lastly, in Appendix A we have found that employing the new FTIR strategy reduces bias with respect to the other measurements by 2 to 3%, which is mostly thanks to the change to HITRAN2020 spectroscopy. Furthermore, the new strategy reduces (at least at Lauder) drifts present in the current FTIR data thanks new regularization resulting in an overall agreement of FTIR with ozonesondes.

Data availability. Current public data for the FTIR, lidar, MWR, and Umkehr can be found at NDACC: https://ndacc.larc.nasa.gov/data. The FTIR data will be updated here with the new strategy explained in the paper. NOAA Dobson Total Column Ozone measurements can also be found on the NOAA GML FTP website here: https://gml.noaa.gov/aftp/ozwv/Dobson/, at WOUDC: https://woudc.org/. Likewise the Monthly Mean optimized/homogenized Umkher profiles can be found on the NOAA GML FTP here: https://gml.noaa.gov/aftp/ozwv/Dobson/AC4/Umkehr/Optimized/. The homogenized Lauder ozonesonde time series are available at the HEGIFTOM ftp-server, with access details described at https://hegiftom.meteo.be/datasets/ozonesondes.

https://doi.org/10.5194/egusphere-2023-2668
Preprint. Discussion started: 20 November 2023

Table A1. Changes from the current FTIR retrieval strategy for ozone to the new strategy. The regularization strength α is specific to the Lauder measurements.

	current strategy	new strategy
Spectroscopy	HITRAN2008	HITRAN2020
Microwindows	1000-1005 cm^{-1}	991.25-993.8 cm^{-1}
		1001.47-1003.04 cm^{-1}
		1005.0-1006.9 cm^{-1}
		1007.35-1009.0 cm^{-1}
Regularization	OEM	Tikhonov with $\alpha = 1000$
A priori	WACCM v6	WACCM IRWG

Appendix A: Influence of FTIR retrieval setup

A1 New FTIR retrieval strategy

The FTIR ozone retrievals used in this study employ an improved retrieval strategy compared to Vigouroux et al. (2008) that has been tested at several NDACC sites and became recently the recommended strategy for the IRWG. The most notable changes to the strategy are listed in Table A1. Firstly, necessary input for atmospheric retrievals is a spectroscopic line list including information on the wavenumber of spectral lines and their line strengths for many molecules. Such information is contained in the spectroscopic database of HITRAN (HIgh resolution TRANsmission), where the IRWG at present uses the HITRAN2008 database (Rothman et al., 2009). In the FTIR retrievals used in this study, we have updated the spectroscopic line list to use the latest HITRAN2020 database (Gordon et al., 2022). We find that a result of this change in spectroscopy is generally that retrieved ozone columns are reduced by $2-3\%$. Secondly, the spectral range that is fitted in the retrieval of ozone (the microwindows) spanned from 1000 to 1005 cm^{-1} in the current strategy. In the new strategy we use a combination of 4 smaller microwindows are chosen such that they avoid strong interference from water vapor lines (García et al., 2022; Schneider and Hase, 2008). Thirdly, the choice of constraints to solve the inverse problem within the retrieval method is chosen as a specific regularization matrix. This matrix can be chosen through the Optimal Estimation Method (OEM, Rodgers, 2000) where the matrix is the inverse of the a priori covariance, or through a so-called smoothing constraint such as the Tikhonov regularization (Tikhonov, 1963). While the first option in theory provides a better regularization from climatological constraints, in practice (such as in the current FTIR retrieval strategy) usually a simplified matrix is adopted to represent the variability on the retrieved profile. In the new FTIR strategy we opt to use the Tikhonov regularization, where the strength of the variability has to be determined by considering the DOFS and the retrieval noise error (Steck, 2002). One important difference with the other regularization method is that only the shape of the profile is constrained within the Tikhonov method and not the absolute value, which reduces trend bias due to a statistic a priori. Lastly, since prior information is important and

https://doi.org/10.5194/egusphere-2023-2668
Preprint. Discussion started: 20 November 2023

Table A2. Drifts with respect to the current FTIR retrieval strategy for each of the measurements in %/decade for every partial column where they have data available. The drifts that are significantly different from zero (considering U_{drift}) are represented in boldface.

	0.5-11 km	14-21 km	21-29 km	29-42 km
Umkehr	-0.8 ± 4.7	$\mathbf{2.7 \pm 1.1}$	$\mathbf{1.9 \pm 1.8}$	$\mathbf{-3.5 \pm 1.1}$
Sonde	$\mathbf{-3.2 \pm 2.1}$	-0.7 ± 1.7	0.5 ± 1.5	–
Lidar	–	0.8 ± 1.9	$\mathbf{3.0 \pm 1.0}$	$\mathbf{-5.2 \pm 1.5}$
MWR	–	–	$\mathbf{2.5 \pm 2.4}$	-1.5 ± 1.9

influential for atmospheric retrievals, we also need to consider the choice of the a priori. This a priori information comes from the Whole Atmosphere Community Climate Model (WACCM, Marsh et al., 2013), where we now adopt a different version named 'IRWG' (Keeble et al., 2021) in stead of version 6 (Gettelman et al., 2019). However, we find no significant effect in comparison to retrievals performed using the current WACCM v6 a priori, but still adopt this change for consistency with the retrieval of other molecules that are targeted by the IRWG which do have significantly different a priori profiles.

A2 Comparison to the current FTIR strategy

To elaborate on the differences between the current FTIR retrieval strategy of Vigouroux et al. (2015) and the new strategy explained in Section 2.1, we showcase here the intercomparison study from above performed with both retrieval strategies. First of all, we consider the bias of the total column of FTIR with Umkehr. The bias of TCO retrieved with the current FTIR strategy is -5.7% with Umkehr, in relative difference as before. This value is higher in absolute value than that found using the new retrieval strategy, where we find a bias of -2.9% with Umkehr. A similar reduction of $1-3\%$ in the bias is also found in most partial columns comparing to the other measurements. Additionally the MAD_s value for the differences in total column between FTIR and Umkehr are also seen to be reduced from 2.0% using the current strategy to 1.7% using the new strategy. This shift in lower ozone columns for the new FTIR strategy is mostly due to the change in spectroscopy from HITRAN2008 to HITRAN2020. Namely, when performing retrievals using the current strategy only changing the spectroscopic data, the columns generally are seen to reduce by 2-3%, which matches the changes in the biases with respect to the other measurements and brings them all in decent agreement with one another considering the instrument uncertainties.

Second of all, the new FTIR retrieval strategy also causes differences in the drifts with respect to the other ground-based instruments. For comparison with earlier results, the drifts derived through the same intercomparison method, but now using the current FTIR strategy, are shown in Table A2. Overall, the change from the current to the new strategy improves most of the drifts and some even change from being significant to being non-significant within the trend error. In the troposphere we find that, while the ozonesonde data does have a drift with respect to current FTIR observations, this drift is no longer present using the new FTIR strategy. In the lower stratosphere, while the drifts do change slightly in value, the conclusions remain the same. In the middle stratosphere the drift with Umkehr is reduced to become non-significant with the new FTIR data. However, one change to remark here is that, while all drifts in the middle stratosphere are reduced, the drift of MWR with the new FTIR

https://doi.org/10.5194/egusphere-2023-2668
Preprint. Discussion started: 20 November 2023

Table B1. Bias and MAD_s values several different time windows where comparison pairs between FTIR and other measurements are constructed. Lidar is omitted here because a window of less than 12h would leave insufficient temporal sampling.

	0.5-11 km M [%], MAD_s [%]	14-21 km M [%], MAD_s [%]	21-29 km M [%], MAD_s [%]	29-42 km M [%], MAD_s [%]
Umkehr 3h	9.7, 16.7	-4.2, 3.5	-7.1, 4.6	- 1.1, 4.7
Umkehr 6h	8.8, 17.1	-4.3, 3.5	-7.4, 4.6	-1.0, 4.7
Umkehr 12h	8.7, 17.5	-4.5, 3.8	-7.4, 4.6	-0.9, 4.7
Sonde 3h	-2.9, 5.1	-6.3, 3.8	-6.7, 3.5	
Sonde 6h	-2.6, 5.4	-6.2, 4.3	-6.8, 3.4	
Sonde 12h	-2.5, 6.0	-6.8, 4.5	-6.8, 3.5	
MWR 1h			-4.9, 5.0	0.3, 6.3
MWR 3h			-5.4, 5.1	-0.1, 6.1
MWR 6h			-5.0, 5.0	-0.2, 6,0

actually increases. This is mainly due to the fact that the time span of the MWR data only lasts until October 2016. The FTIR trend computed on this shorter time series is actually seen to increase when changing from the current to the new strategy, which is the reverse of what happens when the full time series is considered. Potentially then, should the missing 5 years have been included in the MWR observations, following the logic of the other measurements, the drift would have decreased when using the new FTIR and may even no longer be significant. Lastly, in the upper stratosphere, the conclusions remain unchanged when adopting either the current or new FTIR retrieval strategy. Both drifts of Umkehr and lidar with the new FTIR are lower in absolute value than with the current FTIR data.

Previously, we saw that the change in spectroscopy affects the retrieved ozone columns for FTIR, reducing the bias with all other measurements. This effect, however, is the same over the full data set and thus will not alter the derived drifts. The relevant change in the retrieval strategy here is that the regularization in the new strategy uses Tikhonov regularization instead of optimal estimation or that we use a higher regularization strength than in the current strategy bringing down the DOFS. The change in regularization slightly reduces both the positive and the negative trends of FTIR in all partial columns. The consequence we see here is that, when using Tikhonov regularization, the drifts generally improve, and we even find that there is complete agreement of FTIR with ozonesonde over all partial columns.

Appendix B: Choice of time window

The time window where two measurements are used to be compared to one another should not be too large, as to avoid comparing different times of day where ozone could naturally change within the diurnal cycle or by comparing completely different air masses. In part, this is determined by analyzing for which time window there is the lowest bias and scatter between the measurements which are all listed in Table B1. The choice is also made such that there are plenty of comparisons available

https://doi.org/10.5194/egusphere-2023-2668
Preprint. Discussion started: 20 November 2023

to sample the full time coverage and by simultaneously checking the behavior of the drift. The Umkehr observations are only made at sunrise and sunset, so we find that 6 hours is the ideal time window to still have plenty of comparison pairs over the full time series with the best values of M and MAD_s over the partial columns as seen in Table B1. For the same reason, the time window of 6 hours is selected for the comparison to the ozonesonde observations. Because these measurements are not very frequent, we find that we need a large enough window to get a dense enough sampling over the time series. The microwave radiometer measurements are taken a lot more frequently, so a smaller time window of 3 hours here is sufficiently large enough to construct many comparison pairs. The time window of 1h does not really improve the bias or scatter and furthermore makes the bias more negative. Most problematic to construct the comparison pairs however are the lidar observations. These are taken exclusively at night, while FTIR measurements are taken exclusively during the day, because direct sunlight is necessary. A time window of 12 hours is necessary to reach a decent sampling over the available time series.

Appendix C: Effects of smoothing

We explore the difference between the comparisons performed with and without smoothing the high-resolution profile. Both results are shown in Figure C1, where the bias is shown over the full ozone profile together with the MAD_s in the shaded areas. When we perform the intercomparion without smoothing, a lot more oscillation is seen in the relative difference of the profiles. This is especially pronounced for the sonde data, which (as we can also see in Figure 3) shows a lot sharper oscillation than the FTIR profile. These are actual profile measurements from ozonesondes, so direct comparison with FTIR does make sense and limits any influence of trends in the FTIR averaging kernel on the ozonesonde trends (García et al., 2012). However, because the vertical resolution of FTIR is much smaller, these high spatial oscillations can never be observed so it makes sense to adjust the ozonesonde profile to incorporate the same vertical information. We see that this results in overall a better comparison of the profile to FTIR. Similar results, although less pronounced, can be seen for the comparison with the other measurements. When we divide into partial columns, this effect of smoothing on the derived results should not be too big, because we chose the columns such that we have around one DOFS for FTIR. The only exception is in comparing FTIR to Umkehr in the partial columns where the Umkehr DOFS are less than one, where we now use the Umkehr averaging kernels instead. We see for example that the only place where the profile comparison seems to get worse when applying smoothing, is for the Umkehr comparison in the troposphere (where Umkehr reaches a DOFS of 0.5).

To analyze the effects of smoothing on the drift between the instruments, Table C1 shows these values with their uncertainties when smoothing is not applied during the intercomparison. No big differences are seen and the drifts that are deemed significant considering the uncertainty are the same both with and without smoothing. For the ozonesonde comparison, the drift reduces slightly in all partial columns, strengthening the use of a smoothed profile in this comparison. For the other measurements, however, the changes are not as consistent. Both lidar and Umkehr have partial columns where the drift improves after smoothing the profile as well as worsens. For the MWR comparison we see a small increase in the drift in both partial columns. Because the changes are not very large, not changing conclusions about significance within the drift uncertainty, and generally

https://doi.org/10.5194/egusphere-2023-2668
Preprint. Discussion started: 20 November 2023

Table C1. Drifts with respect to FTIR for each of the measurements in %/decade for every partial column where they have data available. The values for the drifts without performing the smoothing step in the intercomparison are shown. The drifts that are significantly different from zero (considering U_{drift}) are represented in boldface.

Drifts, not smooth [%/decade]	Troposphere 0.5-11 km	Lower stratosphere 14-22 km	Middle stratosphere 22-29 km	Upper stratosphere 29-42 km
Umkehr	0.1±4.6	$\mathbf{2.7 \pm 1.4}$	1.1±1.4	$\mathbf{-2.8 \pm 1.1}$
Sonde	0.2 ± 2.5	-1.1 ± 1.4	0.2 ± 1.6	–
Lidar	–	-1.1 ± 1.9	$\mathbf{2.5 \pm 0.9}$	$\mathbf{-3.1 \pm 1.2}$
MWR	–	–	$\mathbf{3.0 \pm 2.0}$	-1.4 ± 2.0

the bias and drifts seem to improve (especially with ozonesonde data) we have chosen to work with a smoothed profile in the intercomparison study.

Appendix D: Effect of discontinuities

In the drift analysis we found that temporal sampling has an important influence on the calculated drift between two measurements. Here we additionally study the influence of potential steps or discontinuities (change point), which for example could arise from physical changes to the measurement instrument. In order to find a change point in the time series of relative differences, we use the Lanzante change-point detection algorithm (Lanzante, 1996) as is similarly done in (García et al., 2014). The Lanzante's algorithm iteratively finds a change point from summing the ranks of the time series from the beginning to each point in the series. Afterwards the series is adjusted using the median of the subseries enclosed by the currently found change points. This method is repeated on the adjusted time series until the found change point is statistically insignificant (p-value<0.05). This method of change-point detection is purely statistical, hopefully aligning with changes to the retrieval strategy or changes to the instrument itself as a cause of the found step.

This method is applied to the relative differences in total and partial columns where we have calculated the drifts in order to find any recurring discontinuities that could point to steps in any of the measurement time series. Multiple change-points over the different intercomparisons are identified using the Lanzante method. One of these steps in the middle of 2018 seems to reappear for multiple measurement comparisons, pointing to a step in the FTIR data (being the reference measurement). To analyze this discontinuity, we show the time series of FTIR DOFS in Figure D1 and again apply the Lanzante change-point detection. Because seasonality is present (which was not the case for the relative differences), we first apply a seasonal fitting to the time series and subtract it. We fit the DOFS $Y(t)$ using

$$Y(t) = A_0 + A_1 t + A_2 \sin(2\pi t) + A_3 \cos(2\pi t) + A_4 \sin(4\pi t) + A_5 \cos(4\pi t)$$

https://doi.org/10.5194/egusphere-2023-2668
Preprint. Discussion started: 20 November 2023

with A_0 the intercept, A_1 the slope and the other terms are the seasonality. From the de-seasoned DOFS we find again the same change point we found for multiple relative differences in May 2018. Since the retrieval strategy is the same over the full time series, we search for the cause of the steps in the instrument itself. From the FTIR instrument logs, we find that on 10 May 2018 there was a 'major alignment' which is the most likely cause for the step in the FTIR data.

Once this change point is found, we can recalculate the drift for each time series of relative differences both before and after this point to find the influence of the discontinuity on the overall drift. Because the step is towards the end of the time series and there are not a lot of data points, the drift to the right of this step has a very high uncertainty. The drift on the left of the step is always similar to the overall drift from Table 4. This means that, for this time series, the step in the FTIR dataset caused by the major alignment does not significantly affect the overall drift reported in this study.

For the other two change points, we cannot identify a clear instrumental reason. Similar to the previous change point, if we calculate the drift left and right of the discontinuities, we find a drifts on the same order as for the total time series, but with larger uncertainties. Such a change-point analysis requires more attention and future work to identify for each instrument the discontinuities and relate them to changes to the instrument or data-processing techniques, but this falls beyond the scope of this paper.

Competing interests. At least one of the (co-)authors is a member of the editorial board of AMT.

Acknowledgements. The work of R. Björklund has been supported by the Belgian Science Policy Office through the IM/RT/23/DORA project.

https://doi.org/10.5194/egusphere-2023-2668
Preprint. Discussion started: 20 November 2023

References

WMO/GAW Report No. 268, 2021: Smit, H. G. J., Thompson, A. M., and ASOPOS panel, Ozonesonde Measurement Principles and Best Operational Practices, ASOPOS (Assessment of Standard Operating Procedures for Ozonesondes) 2.0, WMO Global Atmosphere Watch report series, No. 268, World Meteorological Organization, Geneva.

Ball, W. T., Alsing, J., Mortlock, D. J., Staehelin, J., Haigh, J. D., Peter, T., Tummon, F., Stübi, R., Stenke, A., Anderson, J., Bourassa, A., Davis, S. M., Degenstein, D., Frith, S., Froidevaux, L., Roth, C., Sofieva, V., Wang, R., Wild, J., Yu, P., Ziemke, J. R., and Rozanov, E. V.: Evidence for a continuous decline in lower stratospheric ozone offsetting ozone layer recovery, Atmospheric Chemistry & Physics, 18, 1379–1394, https://doi.org/10.5194/acp-18-1379-2018, 2018.

Basher, R. E.: Review of the Dobson spectrophotometer and its accuracy, in: Atmospheric ozone; Proceedings of the Quadrennial, edited by Zerefos, C. S. and Ghazi, A., p. 387, 1985.

Bass, A. M. and Paur, R. J.: The ultraviolet cross-sections of ozone. I. The measurements. II - Results and temperature dependence, in: Atmospheric ozone; Proceedings of the Quadrennial, edited by Zerefos, C. S. and Ghazi, A., pp. 606–616, 1985.

Bernet, L., Boyd, I., Nedoluha, G., Querel, R., Swart, D., and Hocke, K.: Validation and Trend Analysis of Stratospheric Ozone Data from Ground-Based Observations at Lauder, New Zealand, Remote Sensing, 13, 109, https://doi.org/10.3390/rs13010109, 2020.

Brasseur, G. P. and Solomon, S.: Aeronomy of the Middle Atmosphere: Chemistry and Physics of the Stratosphere and Mesosphere, 2005.

Connor, B. J., Parrish, A., Tsou, J.-J., and McCormick, M. P.: Error analysis for the ground-based microwave ozone measurements during STOIC, Journal of Geophysolical Research, 100, 9283–9291, https://doi.org/10.1029/94JD00413, 1995.

De Mazière, M., Thompson, A. M., Kurylo, M. J., Wild, J. D., Bernhard, G., Blumenstock, T., Braathen, G. O., Hannigan, J. W., Lambert, J.-C., Leblanc, T., McGee, T. J., Nedoluha, G., Petropavlovskikh, I., Seckmeyer, G., Simon, P. C., Steinbrecht, W., and Strahan, S. E.: The Network for the Detection of Atmospheric Composition Change (NDACC): history, status and perspectives, Atmospheric Chemistry & Physics, 18, 4935–4964, https://doi.org/10.5194/acp-18-4935-2018, 2018.

Evans, R. D. and Komhyr, W. D.: Operations Handbook: Ozone Observations with a Dobson Spectrophotometer, vol. 183, World Meteorological Organization Global Atmosphere Watch GAW, 2008.

García, O. E., Schneider, M., Redondas, A., González, Y., Hase, F., Blumenstock, T., and Sepúlveda, E.: Investigating the long-term evolution of subtropical ozone profiles applying ground-based FTIR spectrometry, Atmospheric Measurement Techniques, 5, 2917–2931, https://doi.org/10.5194/amt-5-2917-2012, 2012.

García, O. E., Schneider, M., Hase, F., Blumenstock, T., Sepúlveda, E., and González, Y.: Quality assessment of ozone total column amounts as monitored by ground-based solar absorption spectrometry in the near infrared (> 3000 cm^{-1}), Atmospheric Measurement Techniques, 7, 3071–3084, https://doi.org/10.5194/amt-7-3071-2014, 2014.

García, O. E., Sanromá, E., Schneider, M., Hase, F., Fabián León-Luis, S., Blumenstock, T., Sepúlveda, E., Redondas, A., Carreño, V., Torres, C., and Prats, N.: Improved ozone monitoring by ground-based FTIR spectrometry, Atmospheric Measurement Techniques, 15, 2557–2577, https://doi.org/10.5194/amt-15-2557-2022, 2022.

Gaudel, A., Cooper, O. R., Ancellet, G., Barret, B., Boynard, A., Burrows, J. P., Clerbaux, C., Coheur, P.-F., Cuesta, J., Cuevas, E., Doniki, S., Dufour, G., Ebojie, F., Foret, G., Garcia, O., Granados-Muñoz, M. J., Hannigan, J. W., Hase, F., Hassler, B., Huang, G., Hurtmans, D., Jaffe, D., Jones, N., Kalabokas, P., Kerridge, B., Kulawik, S., Latter, B., Leblanc, T., Le Flochmoën, E., Lin, W., Liu, J., Liu, X., Mahieu, E., McClure-Begley, A., Neu, J. L., Osman, M., Palm, M., Petetin, H., Petropavlovskikh, I., Querel, R., Rahpoe, N., Rozanov, A., Schultz, M. G., Schwab, J., Siddans, R., Smale, D., Steinbacher, M., Tanimoto, H., Tarasick, D. W., Thouret, V., Thompson, A. M.,

https://doi.org/10.5194/egusphere-2023-2668
Preprint. Discussion started: 20 November 2023

Trickl, T., Weatherhead, E., Wespes, C., Worden, H. M., Vigouroux, C., Xu, X., Zeng, G., and Ziemke, J.: Tropospheric Ozone Assessment Report: Present-day distribution and trends of tropospheric ozone relevant to climate and global atmospheric chemistry model evaluation, Elementa: Science of the Anthropocene, 6, 39, https://doi.org/10.1525/elementa.291, 2018.

Geddes, A., Liley, B., McKenzie, R., Kotkamp, M., and Querel, R.: Novel use of a Bentham UV Double Monochromator for measurements of global and direct irradiance, ozone and aerosol, Atmospheric Measurement Techniques Discussions, 2023, 1–17, https://doi.org/10.5194/amt-2023-107, 2023.

Gettelman, A., Mills, M. J., Kinnison, D. E., Garcia, R. R., Smith, A. K., Marsh, D. R., Tilmes, S., Vitt, F., Bardeen, C. G., McInerny, J., Liu, H. L., Solomon, S. C., Polvani, L. M., Emmons, L. K., Lamarque, J. F., Richter, J. H., Glanville, A. S., Bacmeister, J. T., Phillips, A. S., Neale, R. B., Simpson, I. R., DuVivier, A. K., Hodzic, A., and Randel, W. J.: The Whole Atmosphere Community Climate Model Version 6 (WACCM6), Journal of Geophysical Research (Atmospheres), 124, 12,380–12,403, https://doi.org/10.1029/2019JD030943, 2019.

Godin-Beekmann, S., Azouz, N., Sofieva, V. F., Hubert, D., Petropavlovskikh, I., Effertz, P., Ancellet, G., Degenstein, D. A., Zawada, D., Froidevaux, L., Frith, S., Wild, J., Davis, S., Steinbrecht, W., Leblanc, T., Querel, R., Tourpali, K., Damadeo, R., Maillard Barras, E., Stübi, R., Vigouroux, C., Arosio, C., Nedoluha, G., Boyd, I., Van Malderen, R., Mahieu, E., Smale, D., and Sussmann, R.: Updated trends of the stratospheric ozone vertical distribution in the 60° S-60° N latitude range based on the LOTUS regression model, Atmospheric Chemistry & Physics, 22, 11 657–11 673, https://doi.org/10.5194/acp-22-11657-2022, 2022.

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

Gotz, F. W. P., Meetham, A. R., and Dobson, G. M. B.: The Vertical Distribution of Ozone in the Atmosphere, Proceedings of the Royal Society of London Series A, 145, 416–446, https://doi.org/10.1098/rspa.1934.0109, 1934.

Hansen, J., Sato, M., and Ruedy, R.: Radiative forcing and climate response, Journal of Geophysical Research, 102, 6831–6864, https://doi.org/10.1029/96JD03436, 1997.

Jacob, D. J.: Heterogeneous chemistry and tropospheric ozone, Atmospheric Environment, 34, 2131–2159, https://doi.org/10.1016/S1352-2310(99)00462-8, 2000.

Jones, A., Urban, J., Murtagh, D. P., Sanchez, C., Walker, K. A., Livesey, N. J., Froidevaux, L., and Santee, M. L.: Analysis of HCl and ClO time series in the upper stratosphere using satellite data sets, Atmospheric Chemistry & Physics, 11, 5321–5333, https://doi.org/10.5194/acp-11-5321-2011, 2011.

Keckhut, P., McDermid, S., Swart, D., McGee, T., Godin-Beekmann, S., Adriani, A., Barnes, J., Baray, J.-L., Bencherif, H., Claude, H., di Sarra, A. G., Fiocco, G., Hansen, G., Hauchecorne, A., Leblanc, T., Lee, C. H., Pal, S., Megie, G., Nakane, H., Neuber, R., Steinbrecht,

https://doi.org/10.5194/egusphere-2023-2668
Preprint. Discussion started: 20 November 2023

W., and Thayer, J.: Review of ozone and temperature lidar validations performed within the framework of the Network for the Detection of Stratospheric Change, J. Environ. Monit., 6, 721–733, https://doi.org/10.1039/B404256E, 2004.

Keeble, J., Hassler, B., Banerjee, A., Checa-Garcia, R., Chiodo, G., Davis, S., Eyring, V., Griffiths, P. T., Morgenstern, O., Nowack, P., Zeng, G., Zhang, J., Bodeker, G., Burrows, S., Cameron-Smith, P., Cugnet, D., Danek, C., Deushi, M., Horowitz, L. W., Kubin, A., Li, L., Lohmann, G., Michou, M., Mills, M. J., Nabat, P., Olivié, D., Park, S., Seland, Ø., Stoll, J., Wieners, K.-H., and Wu, T.: Evaluating stratospheric ozone and water vapour changes in CMIP6 models from 1850 to 2100, Atmospheric Chemistry & Physics, 21, 5015–5061, https://doi.org/10.5194/acp-21-5015-2021, 2021.

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, Atmospheric Measurement Techniques, 6, 2089–2099, https://doi.org/10.5194/amt-6-2089-2013, 2013.

Langerock, B., De Mazière, M., Hendrick, F., Vigouroux, C., Desmet, F., Dils, B., and Niemeijer, S.: Description of algorithms for co-locating and comparing gridded model data with remote-sensing observations, Geoscientific Model Development, 8, 911–921, https://doi.org/10.5194/gmd-8-911-2015, 2015.

Lanzante, J. R.: Resistant, Robust and Non-Parametric Techniques for the Analysis of Climate Data: Theory and Examples, Including Applications to Historical Radiosonde Station Data, International Journal of Climatology, 16, 1197–1226, https://doi.org/10.1002/(SICI)1097-0088(199611)16:11<1197::AID-JOC89>3.0.CO;2-L, 1996.

Manney, G. L., Santee, M. L., Rex, M., Livesey, N. J., Pitts, M. C., Veefkind, P., Nash, E. R., Wohltmann, I., Lehmann, R., Froidevaux, L., Poole, L. R., Schoeberl, M. R., Haffner, D. P., Davies, J., Dorokhov, V., Gernandt, H., Johnson, B., Kivi, R., Kyrö, E., Larsen, N., Levelt, P. F., Makshtas, A., McElroy, C. T., Nakajima, H., Parrondo, M. C., Tarasick, D. W., von der Gathen, P., Walker, K. A., and Zinoviev, N. S.: Unprecedented Arctic ozone loss in 2011, Nature, 478, 469–475, https://doi.org/10.1038/nature10556, 2011.

Marsh, D. R., Mills, M. J., Kinnison, D. E., Lamarque, J.-F., Calvo, N., and Polvani, L. M.: Climate Change from 1850 to 2005 Simulated in CESM1(WACCM), Journal of Climate, 26, 7372–7391, https://doi.org/10.1175/JCLI-D-12-00558.1, 2013.

McPeters, R. D. and Labow, G. J.: Climatology 2011: An MLS and sonde derived ozone climatology for satellite retrieval algorithms, Journal of Geophysical Research (Atmospheres), 117, D10303, https://doi.org/10.1029/2011JD017006, 2012.

National Research Council: Rethinking the Ozone Problem in Urban and Regional Air Pollution, The National Academies Press, Washington, DC, https://doi.org/10.17226/1889, 1991.

Nedoluha, G. E., Boyd, I. S., Parrish, A., Gomez, R. M., Allen, D. R., Froidevaux, L., Connor, B. J., and Querel, R. R.: Unusual stratospheric ozone anomalies observed in 22 years of measurements from Lauder, New Zealand, Atmospheric Chemistry & Physics, 15, 6817–6826, https://doi.org/10.5194/acp-15-6817-2015, 2015.

Parrish, A.: Millimeter-wave remote sensing of ozone and trace constituents in the stratosphere, Proceedings of the IEEE, 82, 1915–1929, https://doi.org/10.1109/5.338079, 1994.

Parrish, A., Connor, B. J., Tsou, J. J., McDermid, I. S., and Chu, W. P.: Ground-based microwave monitoring of stratospheric ozone, Journal of Geophysical Research, 97, 2541–2546, https://doi.org/10.1029/91JD02914, 1992.

Petropavlovskikh, I., Bhartia, P. K., and DeLuisi, J.: New Umkehr ozone profile retrieval algorithm optimized for climatological studies, Geophysical Research Letters, 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, Atmospheric Measurement Techniques, 15, 1849–1870, https://doi.org/10.5194/amt-15-1849-2022, 2022.

https://doi.org/10.5194/egusphere-2023-2668
Preprint. Discussion started: 20 November 2023

Rinsland, C. P., Connor, B. J., Jones, N. B., Boyd, I., Matthews, W. A., Goldman, A., Murcray, F. J., Murcray, D. G., David, S. J., and Pougatchev, N. S.: Comparison of infrared and Dobson total ozone columns measured from Lauder, New Zealand, Geophysical Research Letters, 23, 1025–1028, https://doi.org/10.1029/96GL00708, 1996.

Rodgers, C. D.: Inverse Methods for Atmospheric Sounding: Theory and Practice, https://doi.org/10.1142/3171, 2000.

Rodgers, C. D. and Connor, B. J.: Intercomparison of remote sounding instruments, Journal of Geophysical Research (Atmospheres), 108, 4116, https://doi.org/10.1029/2002JD002299, 2003.

Rothman, L. S., Gordon, I. E., Barbe, A., Benner, D. C., Bernath, P. F., Birk, M., Boudon, V., Brown, L. R., Campargue, A., Champion, J. P., Chance, K., Coudert, L. H., Dana, V., Devi, V. M., Fally, S., Flaud, J. M., Gamache, R. R., Goldman, A., Jacquemart, D., Kleiner, I., Lacome, N., Lafferty, W. J., Mandin, J. Y., Massie, S. T., Mikhailenko, S. N., Miller, C. E., Moazzen-Ahmadi, N., Naumenko, O. V., Nikitin, A. V., Orphal, J., Perevalov, V. I., Perrin, A., Predoi-Cross, A., Rinsland, C. P., Rotger, M., Šimečková, M., Smith, M. A. H., Sung, K., Tashkun, S. A., Tennyson, J., Toth, R. A., Vandaele, A. C., and Vander Auwera, J.: The HITRAN 2008 molecular spectroscopic database, Journal of Quantitative Spectroscopy and Radiative Transfer, 110, 533–572, https://doi.org/10.1016/j.jqsrt.2009.02.013, 2009.

Santer, B. D., Thorne, P. W., Haimberger, L., Taylor, K. E., Wigley, T. M. L., Lanzante, J. R., Solomon, S., Free, M., Gleckler, P. J., Jones, P. D., Karl, T. R., Klein, S. A., Mears, C., Nychka, D., Schmidt, G. A., Sherwood, S. C., and Wentz, F. J.: Consistency of modelled and observed temperature trends in the tropical troposphere, International Journal of Climatology, 28, 1703–1722, https://doi.org/10.1002/joc.1756, 2008.

Schneider, M. and Hase, F.: Technical Note: Recipe for monitoring of total ozone with a precision of around 1 du applying mid-infrared solar absorption spectra, Atmospheric Chemistry & Physics, 8, 63–71, https://doi.org/10.5194/acp-8-63-2008, 2008.

Schneider, M., Redondas, A., Hase, F., Guirado, C., Blumenstock, T., and Cuevas, E.: Comparison of ground-based Brewer and FTIR total column O_3 monitoring techniques, Atmospheric Chemistry & Physics, 8, 5535–5550, https://doi.org/10.5194/acp-8-5535-2008, 2008.

Smit, H. G. and Oltmans, S. J.: Ozone Data Quality Assessment (O3S-DQA): Resolving inhomogenities due to artifacts in long term ozone sounding records, in: AGU Fall Meeting Abstracts, vol. 2012, pp. A53Q–0443, 2012.

Solomon, S.: Stratospheric ozone depletion: A review of concepts and history, Reviews of Geophysics, 37, 275–316, https://doi.org/10.1029/1999RG900008, 1999.

Solomon, S., Ivy, D. J., Kinnison, D., Mills, M. J., Neely, R. R., and Schmidt, A.: Emergence of healing in the Antarctic ozone layer, Science, 353, 269–274, https://doi.org/10.1126/science.aae0061, 2016.

Steck, T.: Methods for Determining Regularization for Atmospheric Retrieval Problems, Applied Optics, 41, 1788–1797, https://doi.org/10.1364/AO.41.001788, 2002.

Stevenson, D. S., Dentener, F. J., Schultz, M. G., Ellingsen, K., van Noije, T. P. C., Wild, O., Zeng, G., Amann, M., Atherton, C. S., Bell, N., Bergmann, D. J., Bey, I., Butler, T., Cofala, J., Collins, W. J., Derwent, R. G., Doherty, R. M., Drevet, J., Eskes, H. J., Fiore, A. M., Gauss, M., Hauglustaine, D. A., Horowitz, L. W., Isaksen, I. S. A., Krol, M. C., Lamarque, J. F., Lawrence, M. G., Montanaro, V., Müller, J. F., Pitari, G., Prather, M. J., Pyle, J. A., Rast, S., Rodriguez, J. M., Sanderson, M. G., Savage, N. H., Shindell, D. T., Strahan, S. E., Sudo, K., and Szopa, S.: Multimodel ensemble simulations of present-day and near-future tropospheric ozone, Journal of Geophysical Research (Atmospheres), 111, D08301, https://doi.org/10.1029/2005JD006338, 2006.

Tikhonov, A.: Solution of Incorrectly Formulated Problems and the Regularization Method, Soviet Mathematics, 4, 1035–1038, 1963.

Tsou, J. J., Connor, B. J., Parrish, A., Pierce, R. B., Boyd, I. S., Bodeker, G. E., Chu, W. P., Russell, J. M., Swart, D. P. J., and McGee, T. J.: NDSC millimeter wave ozone observations at Lauder, New Zealand, 1992-1998: Improved methodology, validation, and variation study, Journal of Geophysolical Research, 105, 24,263–24,281, https://doi.org/10.1029/2000JD900292, 2000.

https://doi.org/10.5194/egusphere-2023-2668
Preprint. Discussion started: 20 November 2023

Vigouroux, C., de Mazière, M., Demoulin, P., Servais, C., Hase, F., Blumenstock, T., Kramer, I., Schneider, M., Mellqvist, J., Strandberg, A., Velazco, V., Notholt, J., Sussmann, R., Stremme, W., Rockmann, A., Gardiner, T., Coleman, M., and Woods, P.: Evaluation of tropospheric and stratospheric ozone trends over Western Europe from ground-based FTIR network observations, Atmospheric Chemistry & Physics, 8, 6865–6886, https://doi.org/10.5194/acp-8-6865-2008, 2008.

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

Walshaw, C. D.: G. M. B. Dobson - The man and his work, Planetary and Space Science, 37, 1485–1507, https://doi.org/10.1016/0032-0633(89)90141-4, 1989.

WMO: (World Meteorological Organization), Scientific Assessment of Ozone Depletion: 2018, Global Ozone Research and Monitoring Project, Report No. 58, 588 pp., Geneva, Switzerland, 2018.

https://doi.org/10.5194/egusphere-2023-2668
Preprint. Discussion started: 20 November 2023

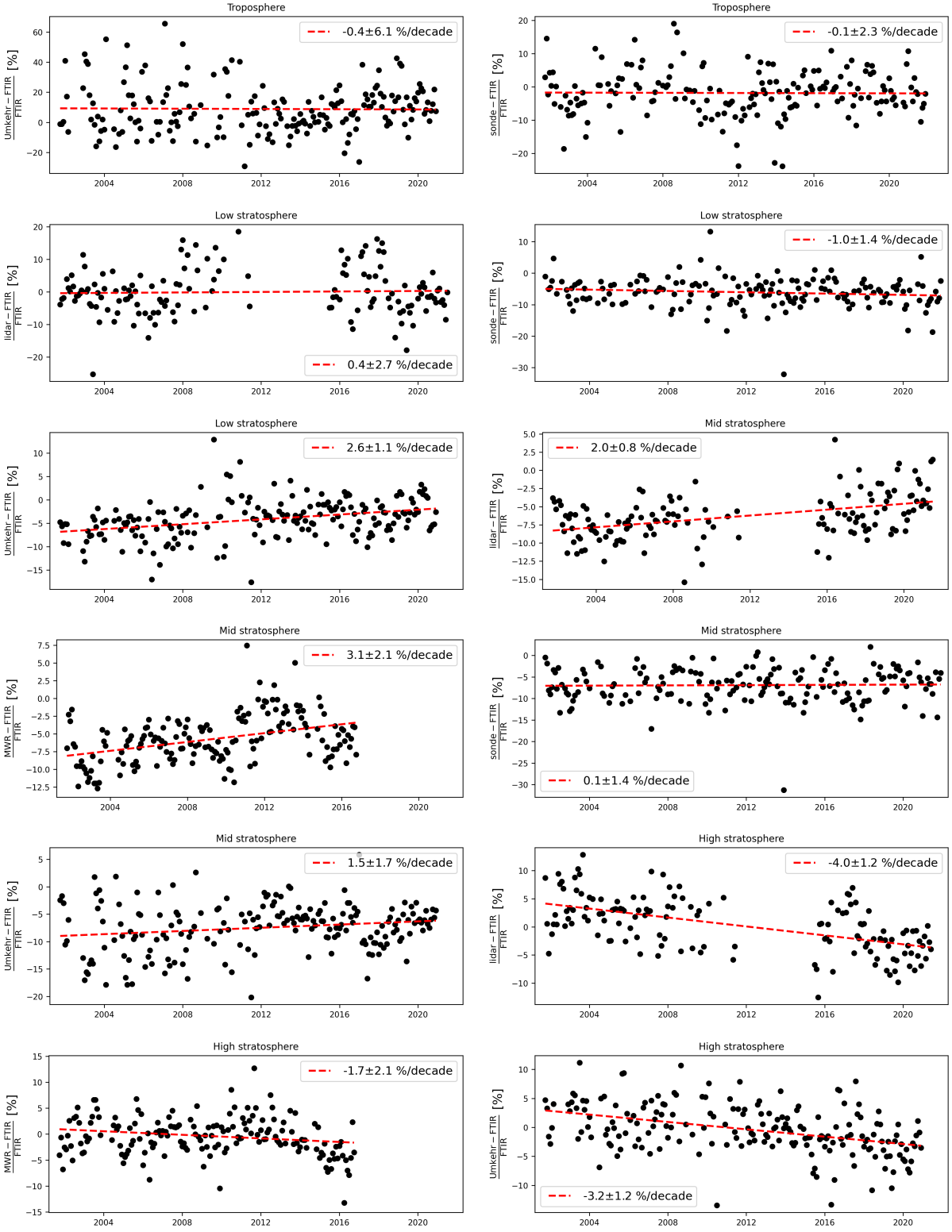

Figure 5. Relative differences between monthly means of sonde, Umkehr, MWR, and lidar with monthly mean of FTIR in the tropospheric and three stratospheric partial columns. Additionally the linear trend fitted to the data is shown with the slope of this linear trend (the drift) and the trend error.

https://doi.org/10.5194/egusphere-2023-2668
Preprint. Discussion started: 20 November 2023

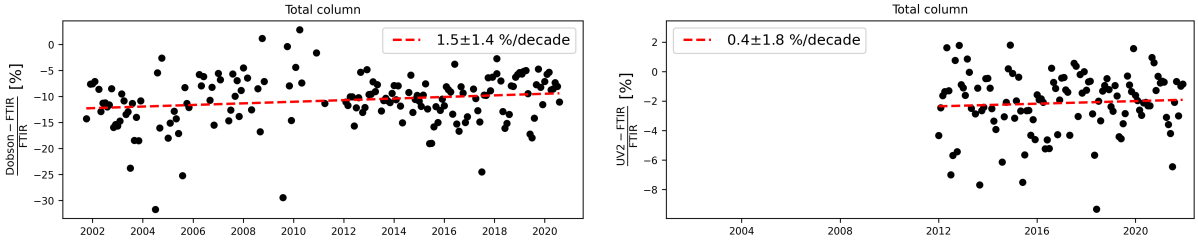

Figure 6. Same as Figure 5 but for total column comparison for Dobson and UV2.

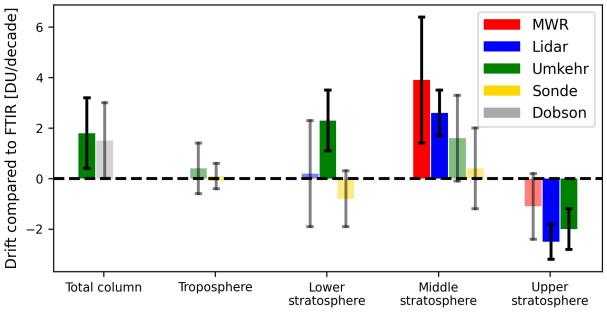

Figure 7. For the total column and each of the partial columns, the drift in DU/decade with respect to FTIR is shown for the Umkehr, ozonesonde, lidar, Dobson, and MWR data. Additionaly, the error margins from the trend analysis are shown. Non-significant trends are shown in a pale color for distinction.

https://doi.org/10.5194/egusphere-2023-2668
Preprint. Discussion started: 20 November 2023

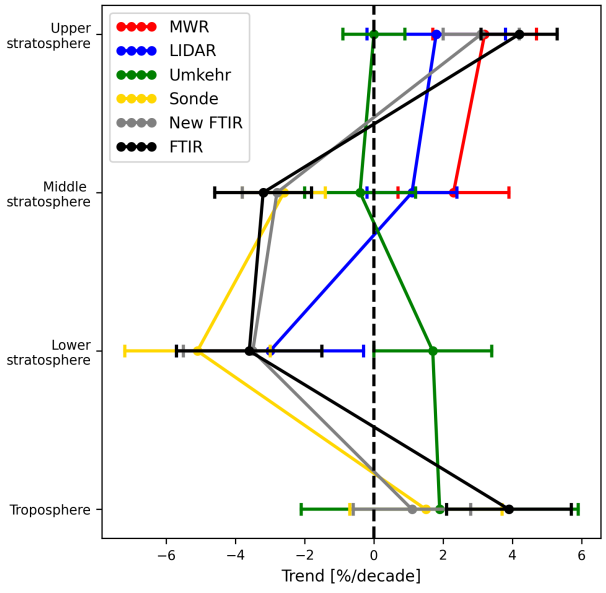

Figure 8. Figure showing the changing trends with altitude of the different measurements for each partial column.

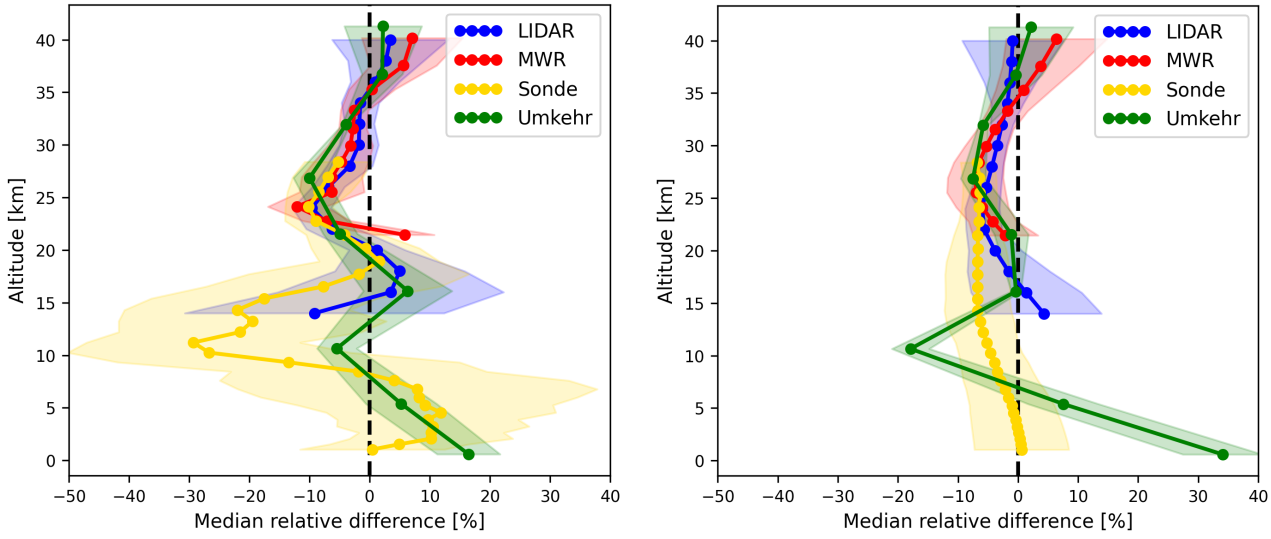

Figure C1. Median bias of the profile when comparing FTIR to lidar, MWR, ozonesonde, and Umkehr together with the MAD_s shown as the coloured, shaded areas. The left panel shows the differences between the profiles when not applying the smoothing step, while the right panel does smooth the higher resolution profile with the averaging kernel of the lower resolution measurement.

https://doi.org/10.5194/egusphere-2023-2668
Preprint. Discussion started: 20 November 2023

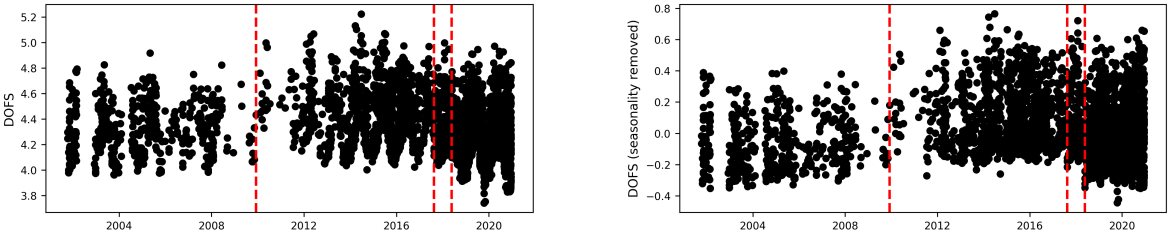

Figure D1. DOFS time series before removing seasonality (upper figure) for the FTIR total column measurements and when removing seasonality as described in the text (lower figure). The statistically significant change points identified by the Lanzante approach are shown in a red, dashed line. The most prominent change point falls around May 2018, with subsequent points found around December 2009 and August 2017.