https://doi.org/10.5194/amt-2021-67
Preprint. Discussion started: 3 May 2021


Atmospheric Measurement Techniques Discussions

# Improved ozone monitoring by ground-based FTIR spectrometry

Omaira E. García[1], Esther Sanromá[1, a], Matthias Schneider[2], Frank Hase[2], Sergio F. León-Luis[1], Thomas Blumenstock[2], Eliezer Sepúlveda[1], Alberto Redondas[1], Virgilio Carreño[1], Carlos Torres[1], and Natalia Prats[1]

[1]Izaña Atmospheric Research Centre (IARC), State Meteorological Agency of Spain (AEMet), Santa Cruz de Tenerife, Spain.
[2]Karlsruhe Institute of Technology (KIT), Institute of Meteorology and Climate Research (IMK-ASF), Karlsruhe, Germany.
[a]Now at: Employment Observatory of the Canary Islands (OBECAN), Santa Cruz de Tenerife, Spain.

**Correspondence:** ogarciar@aemet.es

**Abstract.** Accurate observations of atmospheric ozone ($O_3$) are essential to monitor in detail the key role of $O_3$ in the atmospheric chemistry. The present paper examines the performance of different $O_3$ retrieval strategies from FTIR (Fourier Transform InfraRed) spectrometry by using the 20-year time series of the high-resolution solar spectra acquired from 1999 to 2018 at the subtropical Izaña Observatory (IZO, Spain) within NDACC (Network for the Detection of Atmospheric Composition Change). In particular, the effect of two of the most influential factors have been investigated: the spectral region used for $O_3$ retrievals and inclusion of an atmospheric temperature profile fit. The theoretical and experimental quality assessments of the different FTIR $O_3$ products (total column, TC, amounts and volume mixing ratio, VMR, profiles) provide consistent results. Combining an optimal selection of spectral $O_3$ absorption lines and a simultaneous temperature retrieval results in superior FTIR $O_3$ products, with a precision greater than 0.6-0.7% for $O_3$ TCs as compared to coincident NDACC Brewer observations used as reference. However, this improvement can be only achieved provided the FTIR spectrometer is properly characterised and stable over time. For unstable instruments, the temperature fit has been found to exhibit a strong negative influence on $O_3$ retrievals by increasing the cross-interference between instrumental performance and temperature retrieval. This cross-interference becomes especially noticeable beyond the upper troposphere/lower stratosphere as documented theoretically, as well as experimentally by comparing FTIR $O_3$ profiles to those measured using Electrochemical Concentration Cell (ECC) sondes within NDACC. Consequently, it should be taken into account for the reliable monitoring of $O_3$ vertical distribution, especially on long-term timescales.

## 1 Introduction

Monitoring the atmospheric composition is crucial for understanding the present climate and foreseeing possible future changes and is, therefore, the basis for the design and implementation of efficient climate-change mitigation and adaptation policies. Among the atmospheric gases with important climate effects, ozone ($O_3$) plays a vital role in the atmosphere. In the stratosphere it absorbs a large part of the biologically damaging ultraviolet sunlight, allowing only a small amount to reach the Earth's surface. Likewise, absorption of the ultraviolet radiation heats, stratifies, and determines the vertical stability of the middle atmosphere. In the troposphere, $O_3$ absorbs infrared radiation, acting as an important greenhouse gas, affects the oxidising

https://doi.org/10.5194/amt-2021-67
Preprint. Discussion started: 3 May 2021


Atmospheric Measurement Techniques Discussions

capacity of the atmosphere, and is a harmful phytotoxicant to public health (Cuevas et al., 2013; WMO, 2014a, 2018, and references therein).

$O_3$ measurements have shown a significant decrease in global levels from the 1980s to the 1990s, mainly attributed to the increase of anthropogenic emission of Ozone Depleting Substances (ODSs) during that period (WMO, 2014a, 2018). The implementation of the 1987 Montreal Protocol and its amendments and adjustments has stopped global $O_3$ decay by controlling the ODSs emissions, with $O_3$ concentrations having approximately stabilised since stratospheric ODSs abundances reached their maximum at the end of the 1990s. As a result, global $O_3$ content is expected to slowly increase and return to pre-1980 levels during the 21st century (e.g., Weatherhead et al., 2000; Austin and Butchart, 2003; Eyring et al., 2010; WMO, 2014a, 2018). However, $O_3$ concentrations are not only affected by the presence of ODSs, but also by a wide variety of factors (increase of greenhouse gases concentrations, changes in the Brewer-Dobson circulation and in the stratospheric temperatures, ...), making it very challenging to predict how, when and where the $O_3$ recovery will take place. Hence, high-quality and long-term $O_3$ measurements are essential to further improve our understanding of $O_3$ response to these natural and anthropogenic forcings (Vigouroux et al., 2015).

Within NDACC (Network for the Detection of Atmospheric Composition Change, www.ndaccdemo.org), high-resolution solar absorption infrared spectra have been continuously recorded since the 1990s by ground-based FTIR (Fourier Transform InfraRed) spectrometers distributed at a global scale. By analysing the measured spectra, these instruments are capable of providing both high-quality $O_3$ total column (TC) amounts and low-resolution $O_3$ vertical volume mixing ratio (VMR) profiles (e.g. Barret et al., 2002; Schneider and Hase, 2008; Schneider et al., 2008a; Schneider et al., 2008b; Vigouroux et al., 2008; Viatte et al., 2011; García et al., 2012, 2014; Vigouroux et al., 2015, and references therein). In the last years, the NDACC Infrared Working Group (IRWG, www2.acom.ucar.edu/irwg), and the FTIR community in general, have made considerable efforts in order to standardise as much as possible the inversion strategies used to derive $O_3$ concentrations at the different NDACC stations and, hence, produce uniform and consistent $O_3$ datasets. But, general consensus has not been reached yet and scientific discussion is still on-going. IRWG (2014) summarises the main retrieval recommendations, where many parameters are suggested to be fixed (e.g. the spectral region, the a-priori profiles, and the retrieval grid). However, others are still flexible and station-dependent (e.g. the inclusion of a temperature retrieval or the evaluation of the Instrumental Line Shape, ILS, function), due to either historical reasons or inherent necessities of the different sites (Vigouroux et al., 2008; Vigouroux et al., 2015). Therefore, much efforts should be paid to reach the homogenisation of the FTIR experiments (comprising instruments and retrieval strategies) in order to improve the network-wide consistency and reflect consistent responses to actual variations in atmospheric $O_3$.

In this context, the present paper examines the effect of using different retrieval approaches on the quality of the FTIR $O_3$ products, with the aim of providing an improved $O_3$ strategy that could be applied at any NDACC FTIR station. The influence of two of the most important settings is assessed: the spectral region used for $O_3$ retrievals and simultaneous fit of the atmospheric temperature profile. To our knowledge, so far this analysis has been approached separately in most of the studies present in literature or has not addressed in detail yet. Previous studies have shown, for example, that an optimised selection of the $O_3$ absorption lines or the inclusion of an additional temperature fitting significantly improve the precision of FTIR $O_3$ TCs

https://doi.org/10.5194/amt-2021-67
Preprint. Discussion started: 3 May 2021


Atmospheric Measurement Techniques Discussions

and VMR profiles (e.g., Schneider and Hase, 2008; Schneider et al., 2008b; García et al., 2012, 2014). Nonetheless, possible combined effects were not analysed by these works.

The analysis has been performed at the $O_3$ super-site Izaña Observatory (IZO), where since 1999 ground-based FTIR observations have been carried out coincidentally to other high-quality $O_3$ measurement techniques. By using those data, a comprehensive assessment of the precision and long-term consistency of new $O_3$ retrieval strategies from ground-based FTIR spectrometry can be carried out. To this end, the current paper is structured as follows: Section 2 describes Izaña Observatory, the FTIR measurements and ancillary data considered to assess the quality of the new FTIR $O_3$ products (Brewer TC observations and Electrochemical Concentration Cell, ECC, vertical sondes). Section 3 presents the different FTIR $O_3$ retrieval strategies and their theoretical characterisation in terms of vertical sensitivity and expected uncertainty. Section 4 examines the quality and long-term reliability of the different FTIR $O_3$ TCs and VMR profiles by comparing them to the independent $O_3$ datasets. Finally, Section 5 summarises the main results and conclusions drawn from this work.

## 2 Izaña Observatory and its Ozone Programme

Izaña Observatory is a high-mountain station located on the Island of Tenerife (Spain) in the subtropical North Atlantic Ocean (28.3ºN, 16.5ºW) at an altitude of 2370 m a.s.l.. The observatory is managed by the Izaña Atmospheric Research Centre (IARC, https://izana.aemet.es), which belongs to the State Meteorological Agency of Spain (AEMet, www.aemet.es). IZO is located below the descending branch of the northern subtropical Hadley cell, under a quasi-permanent subsidence regime, and typically above a stable trade wind inversion layer that acts as a natural barrier for local and regional pollution. This strategic location ensures clean air and clear-sky conditions during most of the year, making IZO an excellent station for in-situ and remote-sensing observations (Cuevas et al., 2019, and references therein).

Since many years IZO has run a comprehensive $O_3$ monitoring programme by using different measurement techniques: FTIR, Brewer, and DOAS (Differential Optical Absorption Spectroscopy) spectrometers, as well as ECC $O_3$ sondes and in-situ ultraviolet photometric analysers. The first four techniques routinely contribute to NDACC, aiming at monitoring changes in the troposphere and stratosphere with an emphasis on the long-term evolution of $O_3$ layer, while the in-situ $O_3$ records are taken in the framework of the WMO/GAW (World Meteorological Organization/Global Atmospheric Watch) programme. Refer to Cuevas et al. (2019) for more details about IZO and its atmospheric monitoring programmes.

### 2.1 FTIR Measurements

The IZO FTIR programme has been gathering high-resolution solar spectra within NDACC since 1999, when a Bruker IFS 120M spectrometer was installed by a collaborative agreement between the IMK-ASF (Institute of Meteorology and Climate Research-Atmospheric Trace Gases and Remote Sensing of Karlsruhe Institute of Technology, KIT, www.kit.edu) and AEMet-IARC. In 2005, this instrument was replaced for the upgraded model, the Bruker IFS 120/5HR, which is one of the best performing FTIR spectrometers commercially available. For the present study, the measurements taken from 1999 to 2018 (20 years), encompassing the operation of the two FTIR instruments, have been used.

https://doi.org/10.5194/amt-2021-67
Preprint. Discussion started: 3 May 2021


Atmospheric Measurement Techniques Discussions

Within NDACC activities, the IZO FTIR spectrometer records direct solar absorption spectra in the mid infrared spectral region, i.e., between 740 - 4250 $cm^{-1}$ (corresponding to 13.5 - 2.4 $\mu$m) by using a set of different fieldstops, narrow-bandpass filters, and detectors. Nevertheless, for $O_3$ retrievals, only the 960-1015 $cm^{-1}$ spectral region is considered, which is measured with the NDACC filter 6 using a potassium bromide (KBr) beam splitter and a cooled mercury cadmium telluride (MCT) detector. The solar spectra were taken at a high spectral resolution of 0.0036 $cm^{-1}$ (250 cm maximum optical path difference, OPD, $OPD_{max}$) until April 2000, and at 0.005 $cm^{-1}$ ($OPD_{max}$=180 cm) onward. The IFS 120M's field-of-view (FOV) angle was varied between 0.17° and 0.29° depending on the measurement period, while it was always limited to 0.2° for the IFS 120/5HR. In order to increase the signal-to-noise ratio eight single scans are co-added, thereby the acquisition of one spectrum takes about 10 minutes.

NDACC FTIR solar spectra are only recorded when the line of sight (LOS) between the instrument and the sun is cloud-free. Given the IZO location, the cloud-free conditions are very common and thus FTIR measurements are typically taken about two or three times a week. For the 1999-2018 period the total number of NDACC measurement days for $O_3$ retrievals amounts to 1975, with an annual average of ∼100 measurement days a year. For further details about the FTIR measurements at IZO, refer to Schneider et al. (2005) and García et al. (2012, 2014).

In order to characterise the instrumental performance of the IZO FTIR spectrometers, the ILS function has been routinely monitored about every two months since 1999 using low-pressure $N_2O$-cell measurements and LINEFIT software (v14.5), as detailed in Hase (2012). This ILS treatment ensures the independence of the FTIR trace gas retrievals and instrumental characterisation, but it also allows instrumental alignment and its temporal stability to be verified. Figure 1 depicts the time series of the ILS's modulation efficiency amplitude (MEA) and phase error (PE) parameters, between 1999 and 2018, used for $O_3$ retrievals. Three periods with different features affecting the IZO FTIR measurements can be distinguished: (1) 1999-2004, in spite of $N_2O$-cell measurements were routinely carried out, the ILS estimation is imprecise due to the instability of the IFS 120M spectrometer; (2) 2005-May 2008, the IFS 120/5HR instrument exhibits a gradual temporal drift, but the ILS function is properly assessed; and (3) June 2008-2018, the IFS 120/5HR instrument is optically aligned and well (ILS is nearly nominal). Thereby, these three periods will be independently analysed in the present work in order to examine the influence of instrument status on FTIR $O_3$ products.

### 2.2 Ancillary Data: Brewer and ECC sondes

At IZO Brewer spectrometers, managed by AEMet, have been continuously operating since 1991. In 2001, these activities were accepted by NDACC and, two years later, the RBCC-E (Regional Brewer Calibration Centre for Europe, www.rbcc-e.org) of the WMO/GAW programme was established at IZO. By recording direct solar absorption spectra in ultraviolet spectral region, Brewer instruments provide $O_3$ TCs with an uncertainty better than 1% (Redondas and Cede, 2006). However, recent developments suggest that an overall error of ∼1 Dobson Unit (DU) can be achieved for the IZO RBCC-E reference instruments (Gröbner et al., 2017). The high quality and long-term stability of IZO Brewer observations make them a useful reference for validating ground- and satellite-based instruments (León-Luis et al., 2018).

https://doi.org/10.5194/amt-2021-67
Preprint. Discussion started: 3 May 2021


Atmospheric Measurement Techniques Discussions

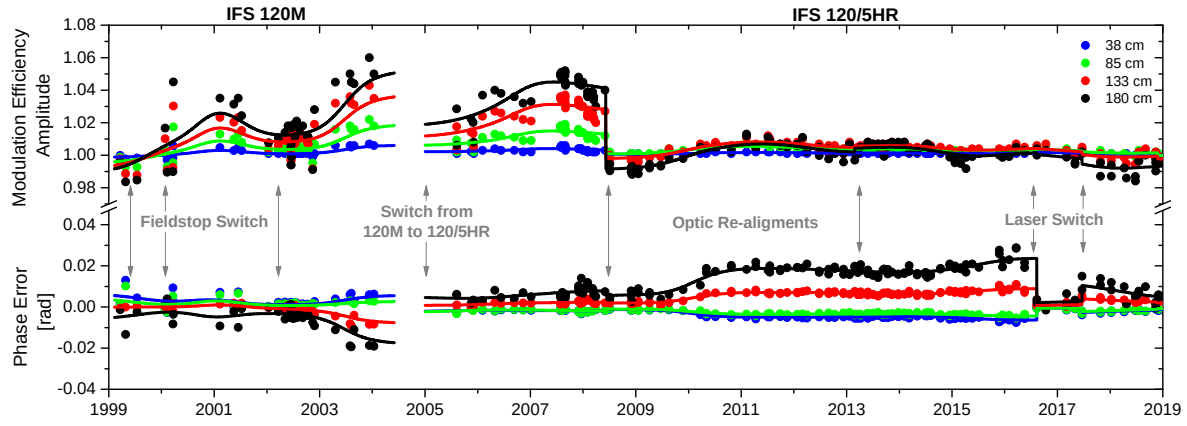



**Figure 1.** Time series of the normalised Modulation Efficiency Amplitude (MEA) and Phase Error (PE, rad) at four optical path differences (33, 85, 133, and 180 cm) for the NDACC $O_3$ measurement settings (filter 6 and MCT detector) of the IZO FTIR spectrometers between 1999 and 2018. Data points represent individual $N_2O$-cell measurements and solid lines depict smoothed MEA and PE curves. Grey solid arrows indicate punctual interventions on the instruments: different changes of fieldstops between 1999 and 2004, switch from the IFS 120M to the IFS 120/5HR system in 2005, optic re-alignments in June 2008 and February 2013, and internal laser replacements in August 2016 and June 2017.

The $O_3$ sonde programme on Tenerife started in November 1992, also run by AEMet, and since March 2001 it has operated in the framework of NDACC. The $O_3$ sounding is based on ECC that senses $O_3$ as it reacts with a dilute solution of potassium iodide (KI) to produce an electrical current proportional to atmospheric $O_3$ concentrations (Komhyr, 1986). The ECC sondes (Scientific Pumps 5A and 6A) were launched once weekly from Santa Cruz de Tenerife station (30 km north-east of IZO, 36 m a.s.l.) until 2010 and, since then, from Botanic observatory (13 km north of IZO, 114 m a.s.l.). The expected uncertainty of the ECC sondes is ±5–15% in the troposphere and ±5% in the stratosphere (WMO, 2014b).

Note that for the purpose of the present work both Brewer and ECC sonde databases fully cover the entire FTIR 1999-2018 period.

## 3 FTIR Ozone Observations

### 3.1 Ozone Retrieval Strategies

To analyse the influence of the spectral region and simultaneous temperature fit on the quality of FTIR $O_3$ products, six different approaches have been defined. They combine three spectral regions and the possibility of performing or not a simultaneous temperature retrieval (referred to as retrieval set-ups 1000, 4MWs, 5MWs, 1000T, 4MWsT and 5MWsT hereafter, Figure 2). Set-up 1000 uses a broad spectral window covering from 1000 to 1005 $cm^{-1}$, which is the one recommended by the NDACC IRWG (IRWG, 2014). This spectral region has been traditionally used by the FTIR community reporting high-quality $O_3$

https://doi.org/10.5194/amt-2021-67
Preprint. Discussion started: 3 May 2021


Atmospheric Measurement Techniques Discussions

products (e.g., Barret et al., 2002; Vigouroux et al., 2008; Lindenmaier et al., 2010; García et al., 2012; Vigouroux et al., 2015). Set-up 5MWs uses five single micro-windows between 991 and 1014 $cm^{-1}$, which is a simplification of the approach suggested by Schneider and Hase (2008). Schneider et al. (2008a) found that this strategy provides more precise $O_3$ estimations than those retrieved from the broad 1000-1005 micro-window when comparing to independent measurements. Set-up 4MWs is the same as 5MWs, but the micro-window at the greatest wavenumbers is discarded in order to avoid any possible saturation of the strong $O_3$ absorption lines contained in this region, especially at high $O_3$ concentrations and low solar elevations. Set-ups 1000T, 4MWsT, and 5MWsT use the same micro-windows as set-ups 1000, 4MWs, and 5MWs respectively, but an optimal estimation of the atmospheric temperature profile is simultaneously carried out. To this end, four $CO_2$ micro-windows are added between 962.80-969.60 $cm^{-1}$ according to Schneider and Hase (2008) (Figure 2).

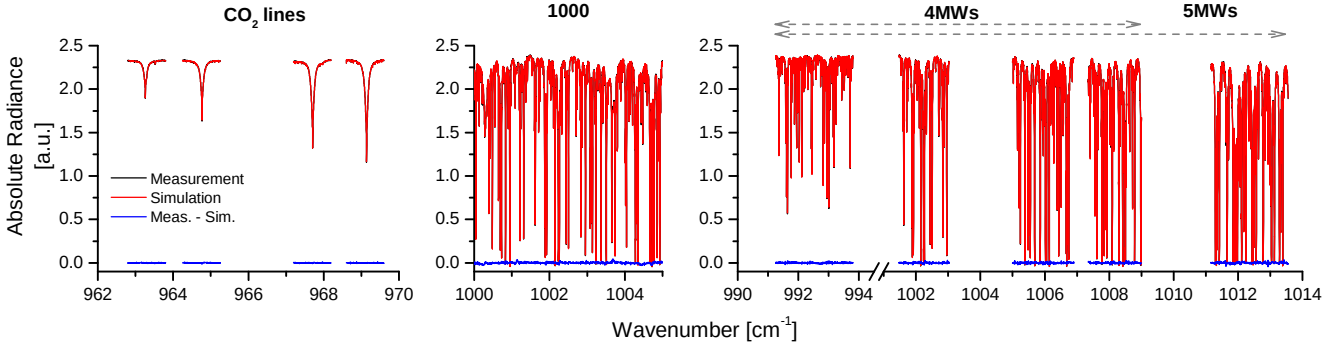



**Figure 2.** Spectral micro-windows used in the different FTIR $O_3$ retrieval strategies. Middle panel shows the broad window used in the set-ups 1000/1000T, encompassing the 1000-1005 $cm^{-1}$ spectral region. Right panel shows the four and five micro-windows used in the set-ups 4MWs/4MWsT, between ∼991 and 1009 $cm^{-1}$, and in 5MWs/5MWsT between ∼991 and 1014 $cm^{-1}$, respectively. Left panel shows the additional micro-windows with four isolated $CO_2$ lines allowing for a simultaneous temperature fit, which are used in the set-ups ending in "T". Red and black lines depict an example of the simulated and measured spectrum taken on $31^{st}$ August 2007 (at solar zenith angle, SZA, of ≈50º, and $O_3$ amount in the slant column, $O_3$ SC, of 390 DU). Blue lines show the difference between the measurement and simulation.

With the exception of the spectral region and temperature treatment, the retrieval strategy is identical for the six approaches. The $O_3$ VMR profiles are derived from the measured solar absorption spectra by means of PROFFIT code (PROFile FIT, Hase et al., 2004), using an ad-hoc Tikhonov-Philips slope constraint (TP1 constraint) on a logarithmic scale. Since $O_3$ concentrations are very variable around the tropopause, the logarithmic inversion has proved to be superior to the linear approach (e.g., Schneider and Hase, 2008; Schneider et al., 2008a; Schneider et al., 2008b). Then, the $O_3$ TCs are computed by integrating the retrieved VMR profiles from the FTIR altitude up to the top of the atmosphere. The remaining settings are based on NDACC IRWG recommendations (IRWG, 2014):

– All set-ups use the same a-priori gas profiles, which are taken from the climatological model WACCM (Whole Atmosphere Community Climate Model, http://waccm.acd.ucar.edu)-version 6 generated by the NCAR (National Center for Atmospheric Research, J. Hannigan, personal communication, 2014).

https://doi.org/10.5194/amt-2021-67
Preprint. Discussion started: 3 May 2021


Atmospheric Measurement Techniques Discussions

- The spectroscopy database used is HITRAN 2008 (Rothman et al., 2009) with a 2009 update for $H_2O$ (www.cfa.harvard.edu).
- The interfering species considered are $H_2O$, $CO_2$, $C_2H_4$, and the main $O_3$ isotopologues ($^{686}O_3$, $^{668}O_3$, $^{676}O_3$, and $^{667}O_3$). In order to minimise interference errors due to $H_2O$, a two-step inversion strategy has been applied: firstly, the actual $H_2O$ profile is derived using a dedicated $H_2O$ retrieval (Schneider et al., 2012) and, then, the $O_3$ retrieval is simultaneously performed with a $H_2O$ scaling fit, which uses the previously derived $H_2O$ state. The remaining interfering species are simultaneously estimated with $O_3$ in this second step. Since the $H_2O$ interference is noticeably reduced by the two-step strategy, this approach is valid for humid FTIR sites (García et al., 2014).
- All set-ups apply the actual ILS time series evaluated from independent $N_2O$-cell measurements (recall Section 2.1).
- The pressure and temperature profiles for forward simulations are taken from the NCEP (National Centers for Environmental Predictions) 12 UT daily database.
- For those approaches performing a simultaneous optimal estimation of atmospheric temperature profile, the NCEP 12 UT daily temperature profiles are used as the a-priori profiles. The a-priori temperature covariance matrix ($\mathbf{S_{aT}}$) has been constructed following to Schneider et al. (2008a).

Once FTIR retrievals are computed, they are filtered according to (1) the number of iterations at which the convergence is reached, and (2) residuals of the simulated–measured spectrum comparison. This ensures that unstable or imprecise observations are not considered (which could likely be introduced, for example, by thin clouds) (García et al., 2016). These two quality flags are applied independently on the six $O_3$ datasets, and only those measurements available for all set-ups are considered in subsequent analysis. This leads to a total of 5393 $O_3$ observations between 1999 and 2018, which are coincident and quality-filtered ($\sim$90% of the original dataset).

### 3.2 Theoretical Quality Assessment

#### 3.2.1 Vertical Sensitivity and Fitting Residuals

Because the vertical resolution of ground-based FTIR measurements is limited, a proper description of the relation between retrieved and true state must be provided together with the retrieved vertical profile. This information is theoretically characterised by the averaging kernel matrix ($\mathbf{A}$) obtained in the retrieval procedure (Rodgers, 2000). The rows of this matrix describe the altitude regions that mainly contribute to the retrieved profile and therefore the vertical distribution of the FTIR sensitivity, while its trace (also so-called "degrees of freedom for signal", DOFS) gives the number of the independent $O_3$ layers detectable by the remote-sensing FTIR instrument. As an example, Figure 3 depicts the $\mathbf{A}$ rows for the 1000 and 5MWs set-ups for the measurement of Figure 2, while Table 1 summarises the DOFS' statistics for the six retrieval strategies considered.

The shape of the $\mathbf{A}$ rows is quite similar for all set-ups with a median total DOFS value of $\sim$4, meaning that the FTIR system is able to roughly resolve four independent atmospheric $O_3$ layers: the troposphere (2.37–13 km), the upper troposphere/lower stratosphere (UTLS) or tropopause region (12–23 km), the middle stratosphere around the ozone maximum (22–29 km), and

https://doi.org/10.5194/amt-2021-67
Preprint. Discussion started: 3 May 2021

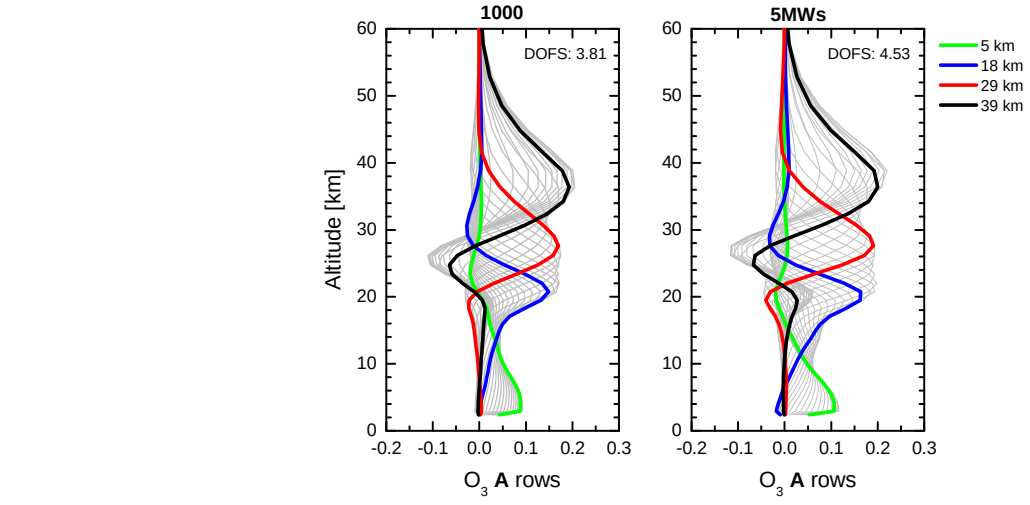




**Figure 3.** Example of averaging kernel (**A**) rows for the set-ups (a) 1000 and (b) 5MWs on a logarithm scale for the measured spectrum of Figure 2. Coloured lines highlight the **A** rows at the altitudes of 5, 18, 29 and 39 km, which are representative for the four layers detectable by the FTIR instrument. Total DOFS values are also shown.

**Table 1.** Summary of statistics of the DOFS and fitting residuals for the set-ups 1000/1000T, 4MWs/4MWsT, and 5MWs/5MWsT for the periods 1999-2004, 2005-May 2008, and June 2008-2018, and for the entire time series (1999-2018). Shown are median (M) and standard deviation ($\sigma$) for each period. The number of quality-filtered measurements is 519, 745, and 4219 for the three periods, respectively, and 5393 for the whole dataset.

| | DOFS | | | | Residuals (x$10^{-3}$) | | | |
|---|---|---|---|---|---|---|---|---|
| | 1999-2004 | 2005-2008 | 2008-2018 | 1999-2018 | 1999-2004 | 2005-2008 | 2008-2018 | 1999-2018 |
| set-up | M, $\sigma$ | M, $\sigma$ | M, $\sigma$ | M, $\sigma$ | M, $\sigma$ | M, $\sigma$ | M, $\sigma$ | M, $\sigma$ |
| 1000 | 3.76, 0.25 | 3.99, 0.21 | 3.98, 0.14 | 3.97, 0.18 | 3.56, 1.90 | 2.62, 0.93 | 2.75, 0.55 | 2.77, 0.93 |
| 4MWs | 4.09, 0.28 | 4.34, 0.12 | 4.30, 0.12 | 4.30, 0.17 | 3.51, 1.90 | 2.57, 0.85 | 2.70, 0.54 | 2.71, 0.91 |
| 5MWs | 4.29, 0.28 | 4.56, 0.15 | 4.52, 0.12 | 4.51, 0.17 | 3.53, 1.90 | 2.58, 0.87 | 2.70, 0.55 | 2.73, 0.92 |
| 1000T | 3.66, 0.31 | 3.98, 0.26 | 3.92, 0.17 | 3.91, 0.22 | 3.50, 1.90 | 2.60, 0.91 | 2.73, 0.55 | 2.75, 0.92 |
| 4MWsT | 3.91, 0.33 | 4.24, 0.18 | 4.17, 0.15 | 4.16, 0.20 | 3.46, 1.90 | 2.55, 0.86 | 2.68, 0.54 | 2.70, 0.91 |
| 5MWsT | 4.10, 0.33 | 4.42, 0.20 | 4.35, 0.15 | 4.35, 0.21 | 3.45, 1.90 | 2.56, 0.87 | 2.68, 0.54 | 2.70, 0.91 |

the upper stratosphere (28–42 km). However, the total DOFS values are found to be greater for those set-ups using narrower micro-windows than for the broad spectral window (see Table 1), whereby the former configurations seem to offer a better vertical sensitivity (especially the 5MWs set-up). This pattern is independent on the FTIR instrument and consistent over time, as observed for the three periods analysed (1999-2004, 2005-May 2008, and June 2008-2018). The comparison among

https://doi.org/10.5194/amt-2021-67
Preprint. Discussion started: 3 May 2021


Atmospheric Measurement Techniques Discussions

periods also reveals, as expected, the lower sensitivity to the $O_3$ concentrations of the IFS 120M spectra as compared to the IFS 120/5HR measurements. The total DOFS values overall differ by 5.5% (1000) and 5.1% (5MWs) between the 1999-2004 and 2008-2018 periods, respectively. When simultaneously fitting the atmospheric temperature profile, the median DOFS values slightly decrease for all strategies, because the information contained in measured spectra is then split into $O_3$ and temperature retrievals (the retrieved state vector space is not perfectly orthogonal). Likewise, the differences among periods become more accentuated (e.g. 6.6% and 5.7% between the 1999-2004 and 2008-2018 periods for 1000T and 5MWsT, respectively). As with DOFS analysis, the fitting residuals are smaller for those set-ups that use narrow micro-windows and apply the temperature fit, allowing a more detailed interpretation of measured spectra to be obtained (as also summarised in Table 1). In addition, the IFS 120M retrievals are found to be considerably more variable than IFS 120/5HR data. However, it is fair to admit that the differences among retrieval strategies lie within the respective error confidence intervals, thereby no robust conclusions can be reached. Note that, in order to make a fair comparison, the fitting residuals are computed as the noise-to-signal ratio for a common spectral region contained in all set-ups (1001.47-1003.04 $cm^{-1}$).

### 3.2.2 Uncertainty Analysis

The characterisation of the different FTIR $O_3$ products has been completed by an uncertainty analysis, which assesses how different error sources could be propagated into the retrieved products. The theoretical error analysis carried out in the present paper is based on Rodgers (2000), and analytically performed by PROFFIT package. The Rodgers' formalism distinguishes three types of error: (1) smoothing error associated with the limited vertical sensitivity of the remote-sensing FTIR instruments, (2) spectral measurement noise, and (3) uncertainties in the input/model parameters (instrumental characteristics, spectroscopy data, ...), which are split into statistical (ST) and systematic (SY) contributions. A detailed description of the uncertainty assessment is given in Appendix A.

In order to assess the effect of $O_3$ absorption signatures on the uncertainty budget, the dependence of the different error sources on $O_3$ slant column (SC) amounts for each retrieval set-up has been examined. This analysis would allow possible inconsistencies among set-ups or saturations of $O_3$ absorption lines at high $O_3$ concentrations to be detected. Figure 4 shows the total and leading error sources (atmospheric temperature profile, ILS, LOS, measurement noise, and spectroscopy) as a function of $O_3$ SC for a typical measurement day at IZO from sunrise to noon. Note that the total error values are calculated as the root-squares-sum of all error sources considered with the exception of the smoothing error. Given that the latter can be considered as an inherent characteristic of the remote-sensing technique, it is not included in the uncertainty assessment suggested by the NDACC IRWG (IRWG, 2014) and, therefore, it has been separately considered in this work.

Consistently, both statistical and systematic uncertainties do dependent on $O_3$ spectroscopic signatures for all set-ups, as observed in Figure 4. When considering statistical error sources, the main contributors are the atmospheric temperature profile for set-ups without temperature fit (error values between 2.0-2.5%), and possible mis-alignments of the FTIR's solar tracker, given by the LOS, for all set-ups (error values up to 1.0%). By contrast, $O_3$ TCs are almost insensitive to errors due to ILS uncertainties and measurement noise (error values smaller than 0.1-0.2%). For the set-ups with a simultaneous temperature retrieval, this fit generates a significant cross-interference with the ILS function, leading to an increment of the ILS error

https://doi.org/10.5194/amt-2021-67
Preprint. Discussion started: 3 May 2021


Atmospheric Measurement Techniques Discussions

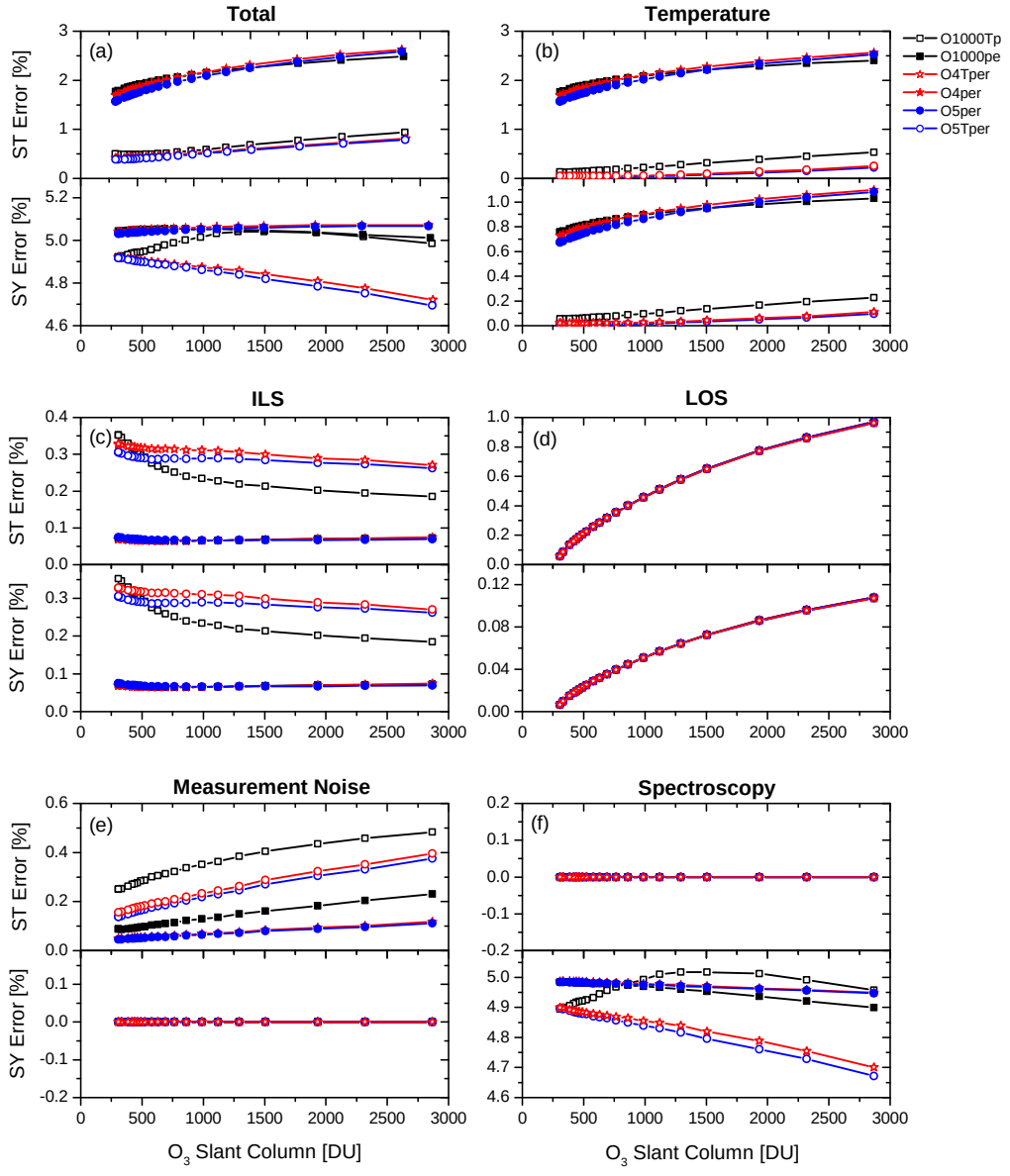



**Figure 4.** Estimated statistical (ST) and systematic (SY) errors [%] for $O_3$ TCs retrieved from the set-ups 1000/1000T, 4MWs/4MWsT, and 5MWs/5MWsT as a function of $O_3$ slant column [DU] for FTIR measurements taken on $31^{st}$ August 2007 from SZAs between 84° (∼07:00 UT) and 21° (∼13:30 UT). (a) Total error, and leading error sources are shown: (b) atmospheric temperature profile, (c) ILS, (d) LOS, (e) spectral measurement noise, and (f) spectroscopy.

https://doi.org/10.5194/amt-2021-67
Preprint. Discussion started: 3 May 2021


Atmospheric Measurement Techniques Discussions

contribution (Schneider and Hase, 2008; García et al., 2012), but also with the measurement noise (especially for 1000T set-up, Figure 4 (e)). However, in return, the temperature error contribution is nearly eliminated, leading to total ST budget considerably improves by ∼1%. Total ST errors range from 0.5 to 1.0% and from 1.5 to 2.5% (between 250 and 3000 DU) for the set-ups with and without simultaneous temperature fit, respectively. Although there are no significant differences among set-ups, the best performance is found in general for the 5MWs/5MWsT strategies.

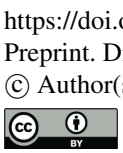
A similar behavior has been documented for the systematic contribution of the leading error sources, where the spectroscopic SY errors determine the total uncertainty budget (with values of ∼5%). A slight decrease of error as $O_3$ SC is increasing can be observed, which becomes more noticeable for set-ups with a simultaneous temperature fit. However, an inconsistency between 4MWsT/5MWsT and 1000T set-ups has been detected. For the 1000T configuration, the spectroscopic SY error exhibits a reverse smile curve with $O_3$ concentrations, and is considerably greater than for the narrow micro-window set-ups. This result might point to a possible saturation of the deeper $O_3$ lines contained in the broad window or some inconsistency in the parameterisation of the temperature dependence of $O_3$ spectroscopy parameters.

The contribution of smoothing error to the total budget is minor for all set-ups without temperature fit independently of the $O_3$ SC range (∼0.2-0.3%). But, when fitting temperature, it gets importance due to the loss of FTIR vertical sensitivity as $O_3$ SC is increasing. This pattern is more significant for the 1000T set-up, with variations from 0.2 to 1.2% versus changes from 0.1 to 1.0% for the 4MWsT/5MWsT set-ups (between 250 and 3000 DU) (data not shown).

Figure 5 illustrates how statistical and systematic uncertainties are vertically distributed for the set-ups 1000/100T and 5MWs/5MWsT for the measurement of Figure 4 at an $O_3$ SC of 390 DU (the 4MWs/4MWsT and 5MWs/5MWsT error profiles are quite similar, thereby the former have been omitted in Figure 5 for simplicity). Total statistical errors smaller than 5% are estimated throughout the atmosphere for all set-ups. As expected, the uncertainty vertical profiles are strongly linked to the atmospheric temperature with maximal errors beyond the UTLS region (where the maximum FTIR sensitivity and the largest $O_3$ concentrations are also located, recall Figure 3). This pattern is consistently observed for both set-ups with and without fitting the atmospheric temperature profiles. However, the error values drastically drop when considering the temperature in the retrieval procedure. Total statistical uncertainties between 1.0-1.5% are expected for the 1000T and 5MWsT set-ups, and as high as 6% when the temperature fit is not taken into account in the retrieval. But, in return, it worsens the ILS contribution especially in the lower/middle stratosphere, making it comparable to the measurement noise error profile. Below these altitudes, the error vertical distributions behave different among set-ups. While statistical uncertainties between 1.0-1.5% are documented throughout the troposphere up to the UTLS region for the 1000/100T and 5MWsT set-ups, the 5MWs duplicates those values in the UTLS (∼3%) and shows larger errors in the lower troposphere (∼2%). In relation to the systematic uncertainty profiles, they range from 5% in the upper troposphere and in the middle stratosphere (around 30 km) up to 8% in the upper stratosphere. As with statistical errors, temperature contribution decreases when including the atmospheric temperature profile in the retrieval procedure, leading to smaller systematic errors in $O_3$ TCs (recall Figure 4).

The error estimation presented here assumes the same set of uncertainty values for all set-ups, which is representative of the IFS 120/HR instrument in the period 2005-2008 (Table A1 in Appendix A). However, some error sources do strongly depend on instrument status (particularly the ILS function, baseline parameters, and solar pointing), affecting the total uncertainty

https://doi.org/10.5194/amt-2021-67
Preprint. Discussion started: 3 May 2021

Atmospheric Measurement Techniques Discussions

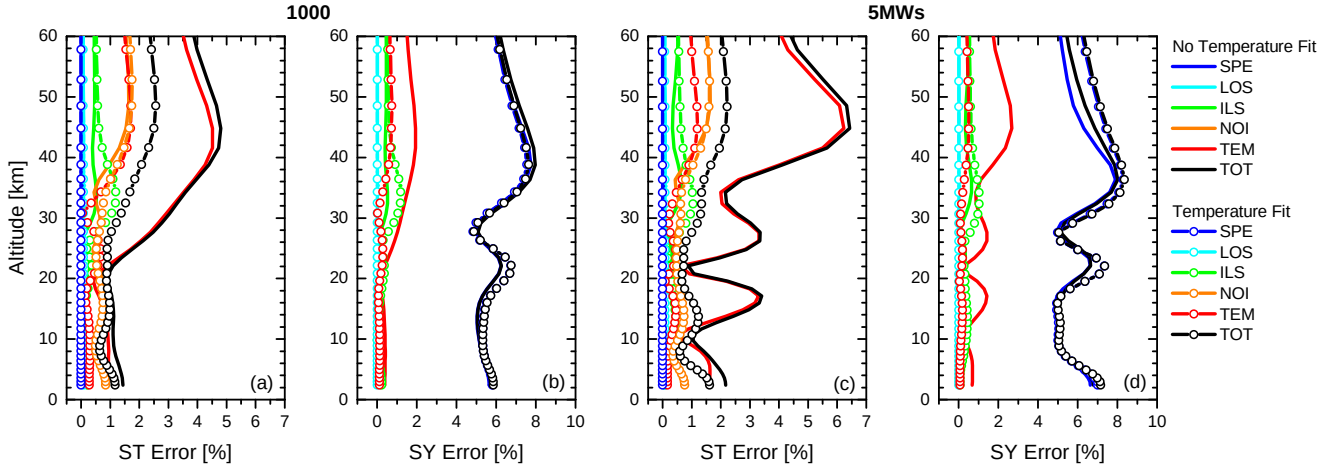



**Figure 5.** Example of estimated statistical (ST) and systematic (SY) error profiles [%] for the 1000/1000T, and 5MWs/5MWsT set-ups for the spectrum of Figure 2. (a) and (b) correspond to ST and SY error profiles, respectively, for the 1000/1000T set-ups. (c) and (d) as for (a) and (b), but for the 5MWs/5MWsT set-ups. Total error (TOT) and leading error sources are shown: spectroscopy (SPE), LOS, ILS, spectral measurement noise (NOI), and atmospheric temperature profile (TEM). Total error profiles are computed as the square root of the quadratic sum of the diagonal elements of the error covariance matrices from statistical and systematic uncertainty sources.

budget. In order to account for the different quality periods of the IZO FTIR instruments, the uncertainty analysis for different set of error values has been included in Appendix A.

To sum up, using several narrow micro-windows instead of a single broad region and applying a temperature profile fit has been found to provide more precise FTIR $O_3$ estimations by increasing the vertical sensitivity and decreasing the expected uncertainties. The simultaneous temperature retrieval could be a suitable approach provided the FTIR system is properly characterised, with a continuous assessment of ILS function, and stable over time (e.g. IFS 120/5HR spectrometers), in order to minimise the negative influence of the ILS uncertainties and measurement noise on $O_3$ retrievals. Finally, although the narrow micro-windows set-ups provide very consistent results, the 5MWsT set-up has theoretically shown to be superior for the typical $O_3$ concentrations observed at the tropical and subtropical latitudes.

## 4 Comparison to Reference Observations

### 4.1 FTIR and Brewer Ozone Total Columns

The performance of the six FTIR $O_3$ retrieval strategies have been assessed by comparing them to coincident Brewer $O_3$ TC data. In order to mitigate the influence of the $O_3$ intra-day variations on comparisons, only FTIR and Brewer measurements

https://doi.org/10.5194/amt-2021-67
Preprint. Discussion started: 3 May 2021


Atmospheric Measurement Techniques Discussions

within a temporal coincidence of 5 minutes have been paired, which makes a total of 2231 coincidences between 1999 and 2018.

Figure 6 displays the time series of Brewer observations together with four examples of FTIR retrievals (for simplicity only the set-ups 1000/1000T and 5MWs/5MWsT are depicted), as well as the time series of corresponding relative differences (RD, FTIR-Brewer). The temporal $O_3$ TCs variations are in general reproduced well by all FTIR products. However, the different performance of the two FTIR instruments becomes evident from this figure: while the RD values of the IFS 120/5HR instrument are very stable over time, the IFS 120M instrument exhibits a more erratic behavior. Besides the greater variability of the IFS 120M and the switch of instrument, the most remarkable feature is a discontinuity detected at the beginning of 2010 by a non-parametric change-point test (Lanzante, 1996) (at 99% confidence level). The systematic jump is of $\sim$1.1% for the set-ups without fitting temperature, and is partially corrected for the set-ups retrieving temperature, $\sim$0.7%. The change-point, already reported by García et al. (2014), is likely due to modifications on the IFS 120/5HR spectrometer (failure of the interferometer's scanner motor and its subsequent replacement). Another change-point was detected by the Lanzante's approach around the beginning of 2014, but of less intensity. It is worth highlighting that both discontinuities were also detected in the differences of FTIR $O_3$ TCs among set-ups, especially among those including the retrieval of the atmospheric temperature. Hence, when no independent observations are available, the analysis of different FTIR products could offer additional tools for identifying inconsistencies and documenting the long-term instrumental stability.

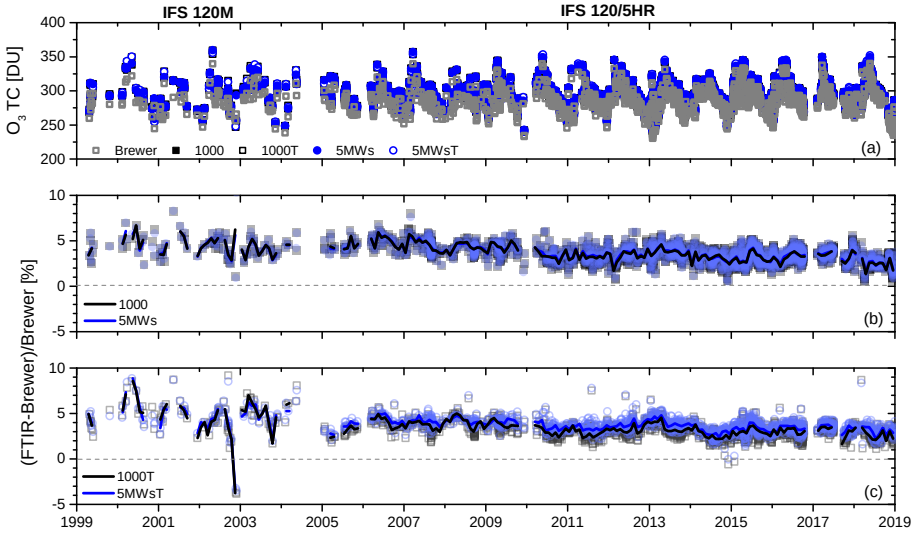



**Figure 6.** Summary of FTIR-Brewer comparison from 1999 to 2018. (a) Time series of $O_3$ TCs [DU] as observed by Brewer and FTIR 1000/1000T, and 5MWs/5MWsT set-ups. (b) Time series of the relative differences RD for the set-ups 1000 and 5MWs, which are calculated as RD[%]=100×($O_3$ $TC_X$ - $O_3$ $TC_Y$)/$O_3$ $TC_Y$, where X and Y means FTIR and Brewer, respectively. (c) As for (b), but for the set-ups 1000T and 5MWsT. Solid lines in (b) and (c) correspond to monthly medians.

https://doi.org/10.5194/amt-2021-67
Preprint. Discussion started: 3 May 2021


Atmospheric Measurement Techniques Discussions

Figure 6 also reveals that, although the scatter of RD is significantly improved by the temperature retrieval, these strategies present in general more extreme values as compared to Brewer data. The RD range between the percentiles $0.1^{th}$ and $99.9^{th}$ is 7.4% and 7.0% for the 1000 and 5MWs set-ups, respectively, while it is 9.3% and 8.6% for the 1000T and 5MWsT configurations, respectively. Note that, for example, the extreme RD values obtained for 1000T and 5MWsT set-ups at the beginning of 2000 and at ending of 2002 (Figure 6 (c)) are not reproduced by the 1000 and 5MWs strategies (Figure 6 (b)). This pattern is consistently observed for all set-ups and over time. The extreme RD values may indicate measurement days with an unusual temperature vertical stratification, which might be wrongly captured by the Brewer and FTIR products assuming a fixed temperature (and pressure) profile. For forward calculations of those FTIR strategies without a simultaneous temperature fit, the temperature and pressure profiles are updated daily from the NCEP database, as previously mentioned, but they are kept constant during the $O_3$ retrieval procedure. Regarding Brewer, no temperature or pressure dependence is considered in the operational data processing (Schneider et al., 2008a; Redondas et al., 2014). Particularly, the Brewer $O_3$ TCs are computed using the so-called effective $O_3$ cross sections throughout the atmosphere (Bass and Paur, 1985), corresponding to an $O_3$ effective height of 22 km and a fixed effective temperature of the $O_3$ layer of -45ºC. These simplifications could produce systematic (seasonal dependence) and random errors (Schneider et al., 2008a; Redondas et al., 2014). In fact, at IZO the effective temperatures significantly differ from the assumed value by Brewer processing in winter months, when the extreme RD values are overall observed (Figure 6 (c)). Nevertheless, a more dedicated study would be desirable to deeply investigate the reasons driving these anomalous values.

When analysing in detail the intercomparison results (Table 2), it is confirmed that the effect of the simultaneous temperature profile retrieval on the FTIR $O_3$ quality depends on the instrumental stability (characterised by the ILS and measurement noise), as expected from the theoretical uncertainty analysis. The agreement between FTIR and Brewer observations significantly gets worse for the more unstable IFS 120M spectrometer when the temperature fit is included in the retrieval procedure (largest median bias and scatter, and least correlation for the 1999-2004 period). Opposite behaviour is documented for the IFS 120/5HR periods: the temperature retrieval consistently improves the precision and accuracy of all FTIR $O_3$ products by reducing considerably the dispersion and bias of the RD distributions. Thus, the best agreement with respect to Brewer data is obtained for set-ups using narrow micro-windows, with a RD scatter of only ∼0.6-0.7% (it increases up to ∼0.8% for the broader region set-up) for the IFS 120/5HR instrument, while it is as high as 1.9% for the IFS 120M when the simultaneous temperature fit is carried out. These values perfectly agree with previous studies (e.g. Schneider et al., 2008a; García et al., 2012), and lie within the expected precision of both instruments (recall Sections 2.2, and 3.2.2, and Appendix A). In fact, as shown in Appendix A, total statistical errors from ∼0.5-1.5% can be expected depending on the instrument status (i.e. the ILS degradation and solar pointing errors). Note that the scatter values of the IFS 120/5HR spectrometer could be a conservative value for the precision of FTIR $O_3$ TC estimations, since they can be interpreted as the root-square-sum of the precision of Brewer and FTIR instruments.

Regarding the systematic differences, a median bias of ∼3-5% is obtained that is mainly attributed to inconsistencies between infrared and ultraviolet spectroscopic parameters. Such differences are consistent with previous studies (e.g., Schneider et al., 2008a; García et al., 2012; García et al., 2016) and with ultraviolet/infrared intercomparison experiments carried out in laboratory (e.g., Piquet-Varrault et al., 2005; Guinet et al., 2010; Gratien et al., 2010). Note that according to the uncertainty

https://doi.org/10.5194/amt-2021-67
Preprint. Discussion started: 3 May 2021


Atmospheric Measurement Techniques Discussions

**Table 2.** Summary of statistics for FTIR-Brewer comparison for the set-ups 1000/1000T, 4MWs/4MWsT, and 5MWs/5MWsT: median (M, in %) and standard deviation ($\sigma$, in %) of the relative differences, and Pearson correlation coefficient (R) of the direct comparison for the periods 1999-2004, 2005-May 2008 and June 2008-2018, and for the entire time series (1999-2018). The number of coincident FTIR-Brewer measurements is 93, 185, and 1892 for the three periods, respectively, and 2170 for the whole dataset.

| | 1999-2004 | 2005-2008 | 2008-2018 | 1999-2018 |
|---|---|---|---|---|
| set-up | M[%], $\sigma$[%], R | M[%], $\sigma$[%], R | M[%], $\sigma$[%], R | M[%], $\sigma$[%], R |
| 1000 | 4.29, 1.38, 0.957 | 4.47, 0.86, 0.970 | 3.35, 0.83, 0.982 | 3.46, 0.95, 0.975 |
| 4MWs | 4.28, 1.36, 0.959 | 4.49, 0.85, 0.9712 | 3.34, 0.82, 0.982 | 3.45, 0.93, 0.976 |
| 5MWs | 4.35, 1.32, 0.962 | 4.53, 0.82, 0.973 | 3.41, 0.81, 0.983 | 3.50, 0.91, 0.977 |
| 1000T | 4.83, 1.97, 0.926 | 3.79, 0.82, 0.972 | 3.04, 0.73, 0.986 | 3.12, 0.90, 0.977 |
| 4MWsT | 4.84, 1.90, 0.934 | 3.97, 0.66, 0.981 | 3.32, 0.68, 0.988 | 3.40, 0.83, 0.981 |
| 5MWsT | 4.81, 1.82, 0.940 | 4.15, 0.63, 0.983 | 3.44, 0.67, 0.988 | 3.53, 0.81, 0.982 |

analysis, errors in $O_3$ spectroscopy could explain a bias in the order of 5% between FTIR and Brewer data. In addition, the bias found slightly depends on the instrument status and temperature retrieval: it is partially corrected by fitting temperature profile for the IFS 120/5HR periods, but the opposite behaviour is observed for the IFS 120M instrument.

Similar conclusions can be reached in general when comparison is performed as a function of $O_3$ signatures in the slant path (Figure 7). The temperature fit improves the performance of the stable instrument and makes it worse for the more unstable instrument, independently of $O_3$ SC range covered at IZO. As with theoretical estimation, it can be seen that the bias between FTIR and Brewer data overall decreases as $O_3$ SCs increase for all set-ups and for IFS 120M and 2005-2008 IFS 120/5HR periods (Figure 7 (a) and (c), respectively). However, for the more stable IFS 120/5HR period (Figure 7 (e)), the bias alters this behaviour for $O_3$ SCs beyond 800 DU, especially for set-ups using narrow micro-windows. Although this pattern is observed independently of including the temperature fit, the differences among set-ups become significant when this simultaneous retrieval is considered. This issue could be attributed to inconsistencies in the parametrisation of the spectroscopic parameters at higher wavenumbers, which get more importance as $O_3$ concentration is increasing. However, the number of FTIR-Brewer coincidences at IZO is rather small for $O_3$ SCs greater than 800 DU (i.e. less than $\sim$5% of Brewer data in the 1999-2018 period), thereby a more robust dataset would be recommendable to better understand what drives this different behaviour. Note that the theoretical systematic inconsistency observed for the 1000T set-up is not confirmed by the experimental comparison. In fact, the 1000/1000T set-ups provide the most accurate $O_3$ TCs with respect to Brewer data for the whole $O_3$ SC range for the IFS 120/5HR periods, but not for IFS 120M. This result further corroborates that the broad region seems to be more sensitive to the negative cross-interference between the temperature retrieval and instrument status.

As pointed out by the theoretical uncertainty analysis, statistical errors are expected to increase with $O_3$ SCs for all the set-ups. This can be seen in the scatter of RD for the IFS 120/5HR periods (Figure 7 (d) and (f)) when the temperature fit is not applied. However, the intercomparison results seem not to exhibit a similar dependence when the temperature retrieval

https://doi.org/10.5194/amt-2021-67
Preprint. Discussion started: 3 May 2021


Atmospheric Measurement Techniques Discussions

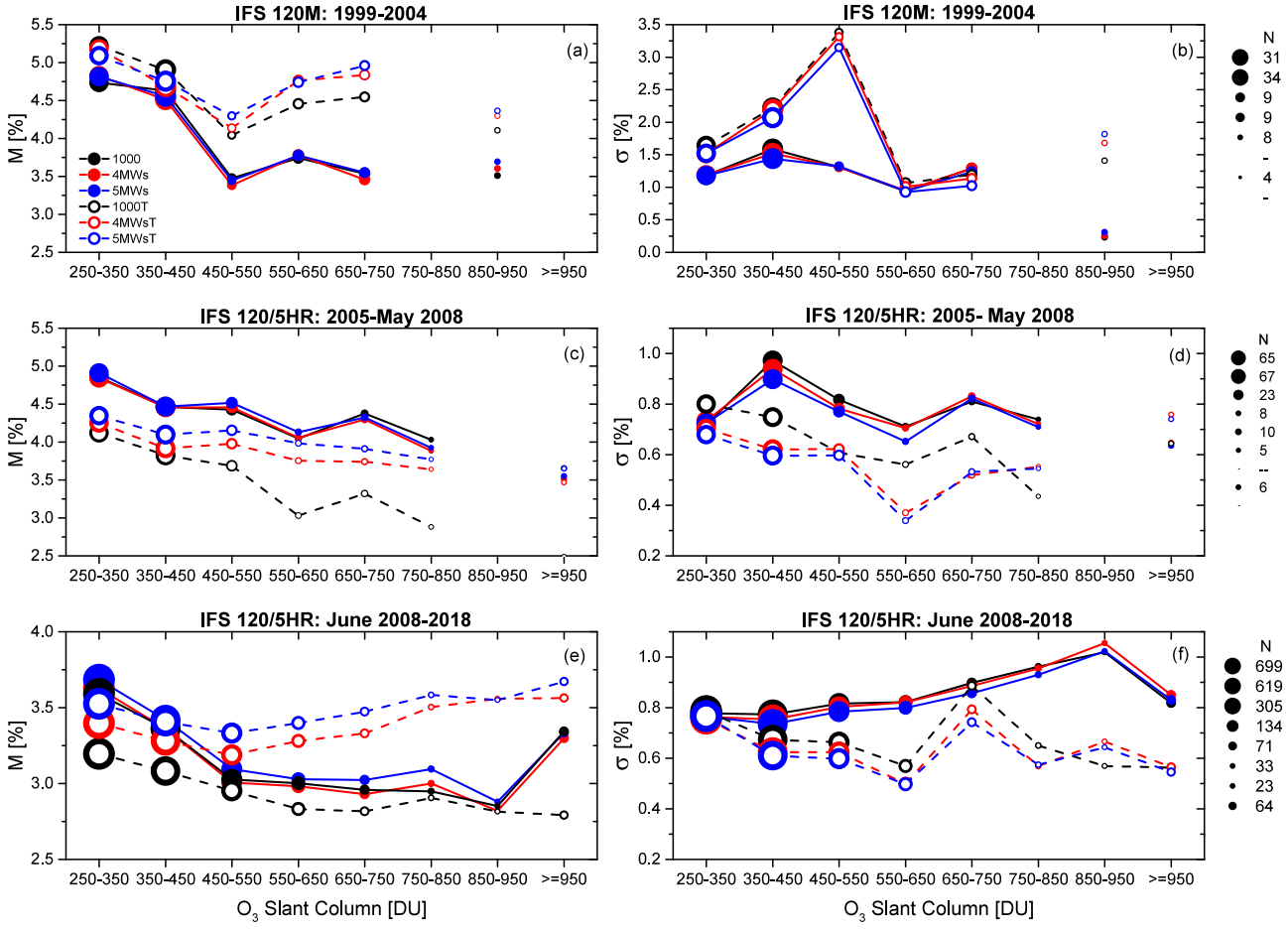
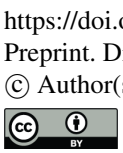



**Figure 7.** Median and standard deviation of relative differences RD (FTIR-Brewer) with respect to Brewer $O_3$ slant column (DU) for the periods 1999-2004, 2005-May 2008, and June 2008-2018. (a), (c), and (e) shows median RD (M, in %) values for the three periods, respectively, and (b), (d), and (f) the same, but for standard deviation of RD ($\sigma$, in %). Dotted area indicates the number of coincident FTIR-Brewer measurements for each $O_3$ SC interval (N), which are included in the legend of each subplot. For a better visualisation a scale factor of 3, 2, and 1 was applied to dotted area for the periods 1999-2004, 2005-May 2008, and June 2008-2018, respectively.

is included, as expected. This could be attributed to an underestimation of the ILS and/or baseline errors in the theoretical assessment: the ILS contribution to total uncertainty budget decreases as $O_3$ SCs increase and becomes more important when the temperature fit is applied (recall Figure 4). The same behaviour is found for baseline error (data not shown). Hence, an increment in the assumed uncertainties for these two error sources, when the temperature takes part of the retrieval procedure, could reconcile partially discrepancies between the theoretical and experiment assessment. Another fact derived from Figure 7 is the possible overestimation of the temperature error contribution. Although a clear improvement between FTIR and Brewer

https://doi.org/10.5194/amt-2021-67
Preprint. Discussion started: 3 May 2021

Atmospheric Measurement Techniques Discussions

data is observed for the whole $O_3$ SC range when the temperature fit is considered, the scatter found is noticeably lower than that predicted when the temperature fit is not considered.

The long-term FTIR time series used in this study allows us to investigate the overall quality and long-term consistency of new products, but also the effects at different timescales. At a seasonal scale, the agreement between FTIR and Brewer is excellent: the annual cycles are completely in phase with Pearson correlation coefficients greater than 0.99 for all retrieval strategies considered. However, the seasonal peak-to-peak amplitudes are slightly different, which produces a seasonal dependence on RD. In order to examine this seasonality, the averaged annual cycles of RD have been computed for the more stable FTIR period (2009-2018, Figure 8). For those approaches without a simultaneous temperature fit (Figure 8 (a)), the RD annual cycle seems to follow the typical $O_3$ TC seasonality at the subtropical latitudes: peak values in spring and minimum in autumn–winter, as a result of the joint effect of the annual shift in the height of subtropical tropopause, and annual cycle of the $O_3$ photochemical production associated to tropical insolation (e.g. García et al., 2014). Hence, a significant correlation between the averaged RD and $O_3$ TC annual cycles for all set-ups is found, with Pearson correlation coefficients ranging from 0.68 to 0.84 for the 1000 and 5MWs strategies, respectively. This may point to Brewer and FTIR products exhibit different sensitivities to $O_3$ seasonal variations.

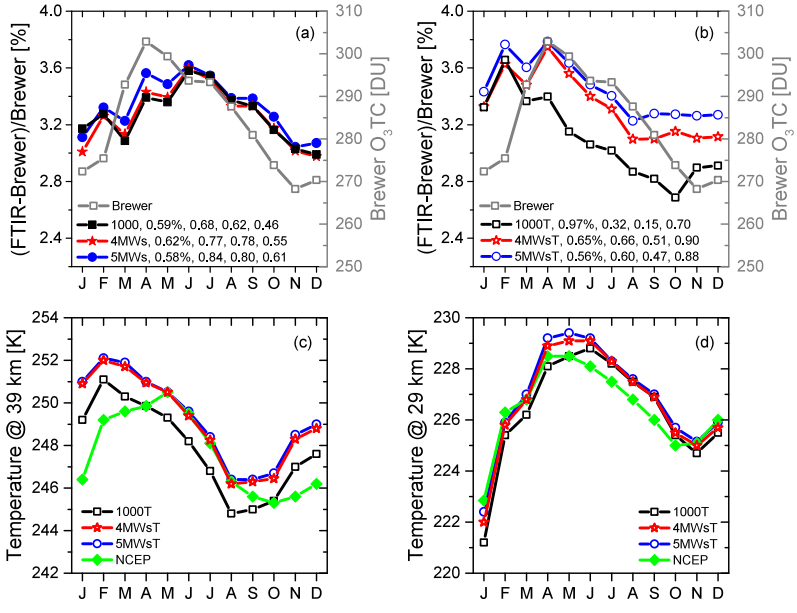



**Figure 8.** (a) Averaged annual cycle of RD for the set-ups 1000, 4MWs, and 5MWs, and for Brewer $O_3$ TCs for the 2009-2018 period. (b) As for (a), but for the set-ups 1000T, 4MWsT, and 5MWsT. (c) and (d) as for (a), but for the atmospheric temperature at 39 km and 29 km, respectively, retrieved from the set-ups 1000T, 4MWsT, and 5MWsT, as well as NCEP database. Legend in (a) and (b) shows the amplitude of RD annual cycle (in %), and Pearson correlation coefficients between RD and Brewer $O_3$ TC annual cycles, and atmospheric temperature (at 29 km and 39 km) annual cycles, respectively.

https://doi.org/10.5194/amt-2021-67
Preprint. Discussion started: 3 May 2021


Atmospheric Measurement Techniques Discussions

The inclusion of the temperature retrieval significantly modifies the RD seasonal pattern (Figure 8 (b)), leading to that the dependence of RD on $O_3$ TCs drops to correlation values between 0.32 and 0.60 for the 1000T and 5MWsT set-ups, respectively. However, in return, a seasonal dependence on the temperature of the upper stratosphere is detected (at 39 km, Figure 8 (c)). The correlation between the averaged annual cycles of the upper stratosphere temperature and RD is ∼0.70 and ∼0.90 for the broad and narrow micro-window set-ups, respectively, including the temperature fit, while it is limited between 0.46 and 0.61 whether the temperature retrieval is not considered. Note that a subtle relationship is found with the temperature in the middle or lower stratosphere (as an example, Figure 8 (d) depicts the averaged temperature annual cycle at 29 km and correlation values are included in the subplots' legend). Additionally, it has been found that the RD seasonal amplitudes are overall augmented by the temperature retrieval. This seasonal artefact could be in part attributed to the different treatment of the atmospheric temperature in the Brewer and FTIR techniques, as stated above for the extreme RD values. The broad spectral region seems to be the most sensitive to this effect: the RD peak-to-peak amplitude goes from 0.59% (1000) to 0.97% (1000T), while it ranges from 0.58% (5MWs) up to 0.56% (5MWsT).

### 4.2 FTIR and ECC Ozone Vertical Profiles

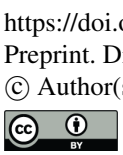
In order to evaluate the influence of the six retrieval strategies on $O_3$ vertical distribution, Figure 9 displays the vertical profiles of the relative differences between FTIR and ECC sondes for the three periods considered, while Table 3 summarises the comparison at the representative altitudes of 5, 18, and 29 km. For this comparison, the approach suggested by Schneider et al. (2008b) and García et al. (2012) was followed, whereby the ECC sondes were corrected daily by comparing them to coincident Brewer data. By means of this correction, the quality and long-term stability of the ECC sonde data can be significantly improved. In addition, the highly-resolved ECC profiles ($\boldsymbol{x_{ECC}}$) were vertically-degraded ($\hat{\boldsymbol{x}}_{\boldsymbol{ECC}}$) applying the averaging kernels obtained in the FTIR $O_3$ retrieval procedure (Rodgers, 2000), as follows:

$$\hat{\boldsymbol{x}}_{\boldsymbol{ECC}} = \mathbf{A}(\boldsymbol{x_{ECC}} - \boldsymbol{x}_a) + \boldsymbol{x}_a, \quad (1)$$

where $\boldsymbol{x}_a$ is the a-priori $O_3$ VMR profiles. By doing this smoothing, on the one hand, the limited sensitivity of FTIR data is taken into account and, on the other hand, the effect of the different strategies on retrieved $O_3$ profiles can be assessed. Note that, in order to homogenise the comparison, only the ECC sondes with continuous measurements up to 29 km have been considered. Beyond this altitude, the ECC data were completed using the a-priori profiles used in FTIR $O_3$ retrievals. Finally, the temporal collocation window between FTIR and ECC sondes is extended to ±3 hours around the sonde launch to ensure sufficient pairs for a robust comparison (N=272 in the 1999-2018 period).

The RD profiles show a strong vertical stratification, whereby the three independent layers detectable by the FTIR systems up to ∼30 km are clearly discernible (troposphere, UTLS, and middle stratosphere, recall Figure 3). Particularly, beyond the UTLS region, the influence of ILS uncertainties on the retrieved $O_3$ profiles is getting importance with altitude, since the full width at half maximum of the narrow $O_3$ absorption lines and ILS function becomes to be comparable. For all set-ups and periods, the simultaneous temperature fit has turned out to worsen the agreement between FTIR and ECC sondes at higher altitudes (see standard deviation profiles in Figure 9 (b), (e), and (h)). However, as the instrument is better aligned and more

https://doi.org/10.5194/amt-2021-67
Preprint. Discussion started: 3 May 2021

Atmospheric Measurement Techniques Discussions

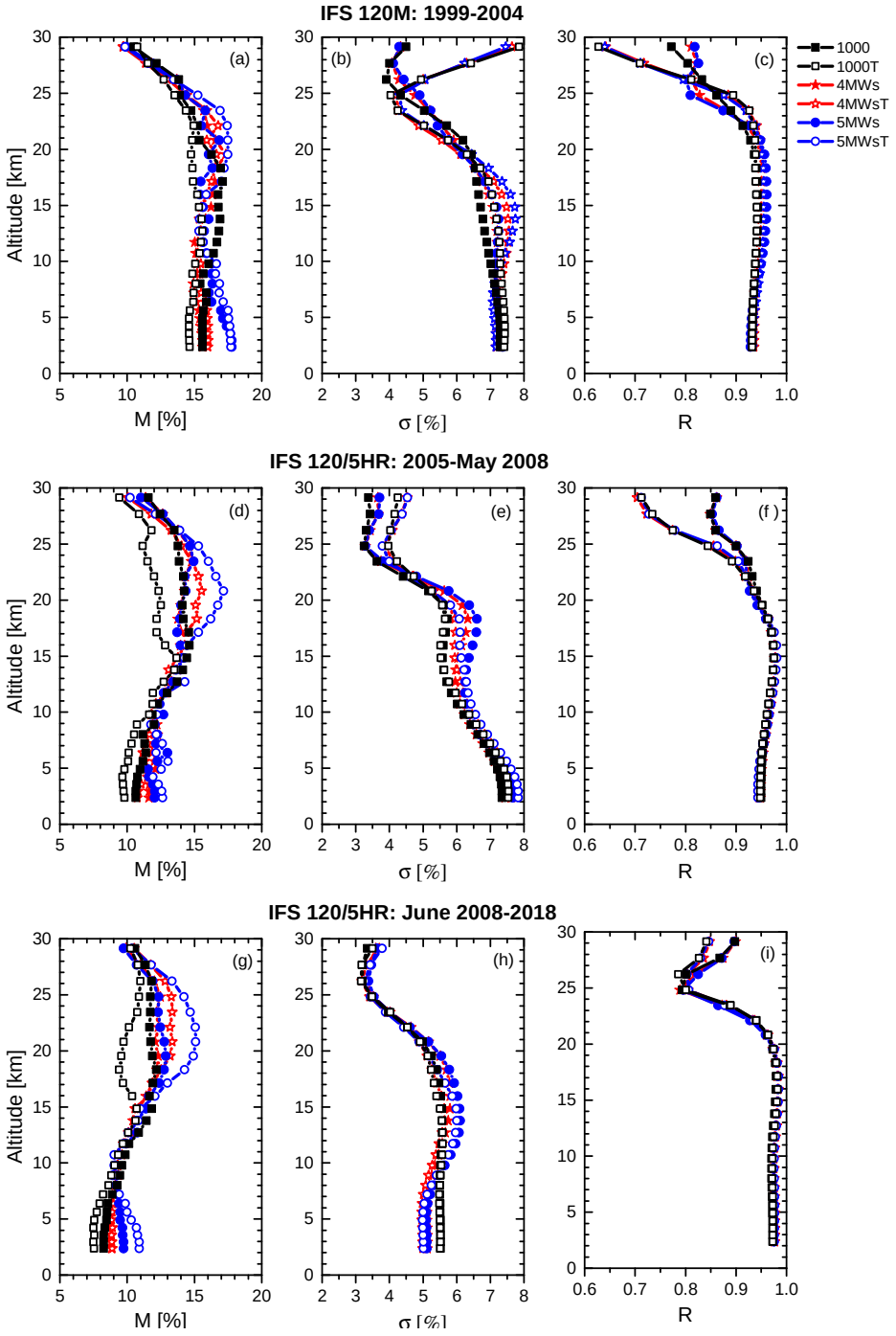



**Figure 9.** Summary of FTIR-smoothed ECC comparison for the periods 1999-2004, 2005-May 2008, and June 2008-2018. (a), (d), and (g) display the vertical profiles of median (M) RD (FTIR-ECC, in %) for the three periods, respectively. (b), (e), and (h) as for (a), (d), and (g), but for standard deviation of RD distributions ($\sigma$, in %). (c), (f), and (i) as for (a), (d), and (g), but for Pearson correlation coefficient. The number of coincident FTIR-ECC measurements is 56, 49, and 167 for the periods 1999-2004, 2005-May 2008, and June 2008-2018, respectively.

https://doi.org/10.5194/amt-2021-67
Preprint. Discussion started: 3 May 2021


Atmospheric Measurement Techniques Discussions

**Table 3.** Same as Table 2, but for FTIR-smoothed ECC comparison at the 5, 18, and 29 km altitude levels. The number of coincident FTIR-ECC measurements is 56, 49, and 167 for the three periods, respectively, and 272 for the whole dataset.

| | 1999-2004 | 2005-2008 | 2008-2018 | 1999-2018 |
|---|---|---|---|---|
| set-up | M[%], $\sigma$[%], R | M[%], $\sigma$[%], R | M[%], $\sigma$[%], R | M[%], $\sigma$[%], R |
| | | FTIR-ECC at 5 km | | |
| 1000 | 15.60, 7.28, 0.933 | 10.97, 7.21, 0.951 | 8.46, 5.52, 0.973 | 10.14, 6.87, 0.952 |
| 4MWs | 15.40, 7.21, 0.935 | 12.12, 7.26, 0.950 | 8.65, 5.12, 0.976 | 10.35, 6.68, 0.952 |
| 5MWs | 17.12, 7.33, 0.931 | 11.55, 7.44, 0.948 | 9.47, 5.10, 0.976 | 10.91, 6.83, 0.953 |
| 1000T | 14.60, 7.43, 0.933 | 9.80, 7.41, 0.949 | 7.55, 5.51, 0.973 | 9.36, 6.93, 0.952 |
| 4MWsT | 15.97, 7.21, 0.936 | 11.53, 7.49, 0.948 | 8.84, 4.97, 0.977 | 10.24, 6.55, 0.956 |
| 5MWsT | 17.55, 7.13, 0.935 | 12.53, 7.62, 0.946 | 10.32, 4.99, 0.976 | 11.40, 6.60, 0.955 |
| | | FTIR-ECC at 18 km | | |
| 1000 | 16.95, 6.55, 0.940 | 14.18, 5.72, 0.963 | 12.17, 5.40, 0.978 | 13.40, 6.02, 0.967 |
| 4MWs | 16.62, 6.55, 0.950 | 13.78, 6.35, 0.961 | 12.20, 5.70, 0.979 | 13.44, 6.29, 0.970 |
| 5MWs | 16.34, 6.55, 0.959 | 13.93, 6.61, 0.959 | 12.74, 5.78, 0.980 | 13.65,6.29, 0.973 |
| 1000T | 14.86, 6.70, 0.939 | 12.19, 5.65, 0.964 | 9.41, 5.24, 0.980 | 11.13, 5.97, 0.969 |
| 4MWsT | 16.27, 7.10, 0.950 | 15.18, 5.95, 0.964 | 12.54, 5.27, 0.982 | 13.89, 5.91, 0.973 |
| 5MWsT | 17.22, 6.95, 0.954 | 16.20, 6.08, 0.966 | 14.27, 5.42, 0.982 | 15.22, 6.04, 0.974 |
| | | FTIR-ECC at 29 km | | |
| 1000 | 10.49, 4.49, 0.772 | 11.58, 3.38, 0.860 | 10.58, 3.34, 0.898 | 10.67, 3.92, 0.850 |
| 4MWs | 9.72, 4.34, 0.811 | 11.37, 3.66, 0.862 | 9.90, 3.60, 0.899 | 9.91, 4.09, 0.860 |
| 5MWs | 9.94, 4.30, 0.818 | 11.02, 3.71, 0.862 | 9.74, 3.63, 0.897 | 9.85, 4.07, 0.861 |
| 1000T | 10.79, 7.86, 0.628 | 9.42, 4.25, 0.713 | 10.25, 3.49, 0.841 | 10.22, 5.27, 0.726 |
| 4MWsT | 9.69, 7.65, 0.642 | 10.06, 4.53, 0.704 | 10.52, 3.70, 0.848 | 10.12, 5.40, 0.740 |
| 5MWsT | 9.84, 7.46, 0.640 | 10.22, 4.54, 0.719 | 10.38, 3.79, 0.846 | 10.29, 5.41, 0.741 |

https://doi.org/10.5194/amt-2021-67
Preprint. Discussion started: 3 May 2021


Atmospheric Measurement Techniques Discussions

stable over time, the effect of this cross-interference becomes less significant until no noticeable differences are observed for the 2008-2018 period. As shown in Table 3, the scatter at 29 km for FTIR-ECC comparison only changes from 3.6% to 3.8% for the 5MWs and 5MWsT set-ups, respectively, for the 2008-2018 period, while the variation ranges between 4.3% and 7.5% for the 1999-2004 period. In the UTLS region (18 km in Table 3), the performance is improved by the temperature fit for all set-ups and for the IFS 120/5HR periods, but not for the IFS 120M. This may indicate the prevalence of negative effect of the ILS uncertainties and measurement noise over the improvement attributed to the temperature fit. The same pattern is documented for the tropospheric $O_3$ concentrations even though the differences among retrieval strategies are not as significant as at higher altitudes. Note that these scatter values agree well with the expected uncertainty for ECC sondes (∼5–15%) and with the FTIR theoretical error estimation (recall Section 3.2.2), as well as with previous works (Schneider et al., 2008b; García et al., 2012; Duflot et al., 2017, and references therein). As stated in these studies, the limited vertical sensitivity of the FTIR profiles could account for part of the dispersion observed between both datasets. Other sources of discrepancies might be the different observing geometries (i.e. the two measurement techniques sample different air masses).

The vertical stratification observed in the median RD profiles also differs between both FTIR instruments (Figure 9 (a), (d), and (g)). While all set-ups consistently show larger biases up to the UTLS region for the IFS 120M as compared to the IFS 120/5HR, the differences between instruments are minimised beyond the middle stratosphere. Although the influence of temperature retrieval is more subtle for the IFS 120M period, indicating a compensation of the temperature, measurement noise, and ILS uncertainties, its effect is significant for the IFS 120/5HR periods. For the latter, the 1000T set-up improves its accuracy beyond the UTLS region by including the temperature retrieval, while the narrow micro-windows set-ups are overall found to worsen for these layers. The theoretical uncertainty estimation pointed out this differentiated behaviour, but with less intensity (recall Figure 5).

In summary, considering the 2008-2018 period as reference (better instrumental alignment and more FTIR-ECC coincidences), the best performance is overall documented for the set-ups using narrow micro-windows and simultaneous temperature fit up to the upper troposphere region. Beyond these altitudes, the broad micro-window strategy seems to provide the better agreement with respect to ECC data. Nevertheless, the differences among strategies lie within the respective error confidence intervals, thereby no robust conclusions can be reached. In addition, it is fair to admit that the ECC comparison only allows the FTIR vertical profiles to be analysed in detail up to ∼30 km. However, compensations among the ILS, temperature, measurement errors, and $O_3$ vertical distribution should occur at higher altitudes, leading to the usage of narrow micro-windows (and temperature fit) clearly provide the best results in the integrated total columns, as documented by the FTIR-Brewer comparison. Unfortunately, ECC sondes do not usually reach altitudes higher than 30-34 km, thereby other measurement techniques, such as microwave or LIDAR $O_3$ profiles, would be of great use to further complete the quality assessment.

https://doi.org/10.5194/amt-2021-67
Preprint. Discussion started: 3 May 2021


Atmospheric Measurement Techniques Discussions

## 5 Summary and Conclusions

Accurate ozone ($O_3$) products are mandatory to monitor the evolution of the Earth's atmosphere system. In this context, the current paper has assessed the effect of using different retrieval strategies on the quality of $O_3$ products from ground-based NDACC FTIR spectrometry, with the aim of providing an improved $O_3$ retrieval that could be applied at any NDACC FTIR station. For this purpose, the high-quality NDACC FTIR measurements taken at the subtropical Izaña Observatory (IZO) between 1999 and 2018 has been utilised. The 20-year time series of $O_3$ observations has allowed us to assess, on the one hand, the quality and long-term consistency of the different FTIR $O_3$ products and, on the other hand, to evaluate the influence of instrumental status on the $O_3$ retrievals.

Quality of the FTIR $O_3$ products improves as the retrieval strategies become more refined by including $O_3$ absorption lines in specific narrow micro-windows (between 991 and 1014 $cm^{-1}$) instead of using a broad spectral region (between 1000-1005 $cm^{-1}$). Approaches using narrow micro-windows have theoretically and experimentally proven to be superior by providing greater vertical sensitivity, smaller expected uncertainties, and better agreement with respect to independent data. The optimal selection of the spectral $O_3$ micro-windows can enhance the precision of FTIR $O_3$ TCs by $\sim$0.1-0.2% with respect to coincident Brewer observations used as reference. But, at the same time, they have shown to be consistent with the standard NDACC set-up (i.e. no important biases were found among the different retrieval strategies). In addition, independently of the $O_3$ absorption lines used, the simultaneous atmospheric temperature retrieval has been found to be a very useful tool for $O_3$ monitoring by ground-based NDACC FTIR systems. Theoretically, the total error of $O_3$ TCs is halved when applying a temperature fit and, experimentally, the scatter with respect to Brewer data is found to be reduced up to $\sim$0.2% for those strategies also using narrow $O_3$ absorption lines. However, this improvement can be only reached provided the FTIR instrument is properly characterised and stable over time (e.g. IFS 120/5HR spectrometer). For more unstable instruments, such as the IFS 120M, the inclusion of atmospheric temperature fit in the $O_3$ retrieval procedure may not be recommendable, since it worsens the quality of FTIR $O_3$ products by increasing the cross-interference with instrumental performance.

The effect of the most influential settings on FTIR $O_3$ retrieval procedure has been examined in this paper. Nevertheless, there is great potential to further improve the precision and accuracy of FTIR $O_3$ products, as well as their harmonisation within the NDACC community (i.e. with the treatment of the instrumental response). An improved $O_3$ monitoring could help to estimate more precisely the small expected signal of recovery or decline of $O_3$ concentrations, both for integrated total columns and vertical distributions, at a global scale. This is particularly challenging in those regions where $O_3$ concentrations are less variable, such as at tropical and subtropical latitudes. Furthermore, new opportunities would show up to better understand the different and outstanding $O_3$ impacts on Earth climate system, improving their representation in the current global climate models and, thus, the knowledge of their long-term evolution.

https://doi.org/10.5194/amt-2021-67
Preprint. Discussion started: 3 May 2021


Atmospheric Measurement Techniques Discussions

## Appendix A: Uncertainty Budget

Theoretical uncertainties of FTIR products can be estimated by following the formalism detailed by Rodgers (2000), which includes the effect of smoothing error, spectral measurement noise, and different model parameter sources. The difference between the retrieved state, $\hat{\boldsymbol{x}}$, and real state, $\boldsymbol{x}$, can therefore be written as a linear combination of the a-priori state, $\boldsymbol{x}_a$, real and estimated model parameters, $\boldsymbol{p}$ and $\hat{\boldsymbol{p}}$ respectively, and measurement noise $\boldsymbol{\epsilon}$:

$$(\hat{\boldsymbol{x}} - \boldsymbol{x}) = (\mathbf{A} - \mathbf{I})(\boldsymbol{x} - \boldsymbol{x}_a) + \mathbf{G}\mathbf{K}_p(\boldsymbol{p} - \hat{\boldsymbol{p}}) + \mathbf{G}\boldsymbol{\epsilon}, \tag{A1}$$

where $\mathbf{G}$ represents the gain matrix, $\mathbf{K}_p$ the sensitivity matrix to the model parameters, $\mathbf{I}$ the identify matrix, and $\mathbf{A}$ the averaging kernel matrix.

The first term of Eq. (A1) refers to smoothing error, which has been calculated as $(\mathbf{A}-\mathbf{I})\mathbf{S}_{\mathbf{aO_3}}(\mathbf{A}-\mathbf{I})^T$. The $\mathbf{S}_{\mathbf{aO_3}}$ matrix is the $O_3$ a-priori covariance matrix, which has been computed in this work from the 1999-2018 ECC sonde climatology at IZO according to Schneider and Hase (2008). Note that ECC sondes usually burst between 30 and 34 km, hence this climatology was completed beyond 31 km by using the WACCM-version 6 simulations for subtropical latitudes.

The error covariance matrix for measurement noise ($\mathbf{S}_{\mathbf{x},\epsilon}$) is analytically calculated by

$$\mathbf{S}_{\mathbf{x},\epsilon} = \mathbf{G}\mathbf{S}_{\mathbf{y},\epsilon}\mathbf{G}^T, \tag{A2}$$

where $\mathbf{S}_{\mathbf{y},\epsilon}$ is the covariance matrix for measurement noise in the measurement.

The error contribution of the model parameters $\boldsymbol{p}$ can be analytically estimated through the respective error covariance matrices $\mathbf{S}_{\mathbf{x},\mathbf{p}}$:

$$\mathbf{S}_{\mathbf{x},\mathbf{p}} = \mathbf{G}\mathbf{K}_\mathbf{p}\mathbf{S}_\mathbf{p}{\mathbf{K}_\mathbf{p}}^T\mathbf{G}^T, \tag{A3}$$

where $\mathbf{S}_\mathbf{p}$ is the covariance matrix of the uncertainties $\mathbf{\Delta p}$. In the current paper, $\mathbf{S}_\mathbf{p}$ is estimated considering error sources, values, and partitioning between random and systematic contributions listed in Table A1. They have been identified as the typical error sources and values affecting the different FTIR products (e.g., Hase, 2007; Schneider and Hase, 2008; García et al., 2016, and references therein). Total statistical and systematic uncertainties (displayed in Figure 4 and 5) are then calculated as the square root sum of the squares of all statistical and systematic errors considered, respectively. Note that measurement noise is considered as purely random, while spectroscopic parameters are purely systematic.

Table A1 lists the uncertainty values representative of the IFS 120/HR instrument in the period 2005-2008. However, the error sources associated to instrument status (i.e. ILS function, baseline parameters, and solar pointing) can be different depending on the quality periods of the IZO FTIR spectrometers. In order to account for this fact, Figure A1 summarises the effect of the different sets of error values on $O_3$ TCs for the measurement day of Figure 4. Note that this figure only includes the error estimations for different ILS and LOS configurations (due to the influence of baseline on total error budget is of minor importance), and for the set-ups 1000/1000T and 5MWs/5MWsT set-ups (the 4MWs/4MWsT error estimations are quite similar to the 5MWs/5MWsT ones, thereby they have been omitted for simplicity).

https://doi.org/10.5194/amt-2021-67
Preprint. Discussion started: 3 May 2021


Atmospheric Measurement Techniques Discussions

**Table A1.** Error sources and assumed values used in the theoretical uncertainty estimation. Last column shows the contribution of statistical (ST) and systematic (SY) sources to total error. Chann.: Channeling; MEA: Modulation Efficiency Amplitude; PE: Phase Error; Int.: Intensity; $\nu$-scale: Spectral position; S: Intensity parameter; $\nu$: Pressure broadening parameter.

| Error Source | Error | ST/SY |
|---|---|---|
| Baseline (Chann. and Offset) | 0.1% and 0.1% | 50/50 |
| MEA and PE (ILS) | 1% and 0.01 rad | 50/50 |
| Temperature profile | 2 K (<50 km) | 70/30 |
| | 5 K (>50 km) | 70/30 |
| Line of Sight (LOS) | 0.001 rad | 90/10 |
| Solar Lines (Int. and $\nu$-scale) | 1% and $10^{-6}$ | 80/20 |
| Spectroscopy | 5% for S and 5% $\nu$ | 0/100 |

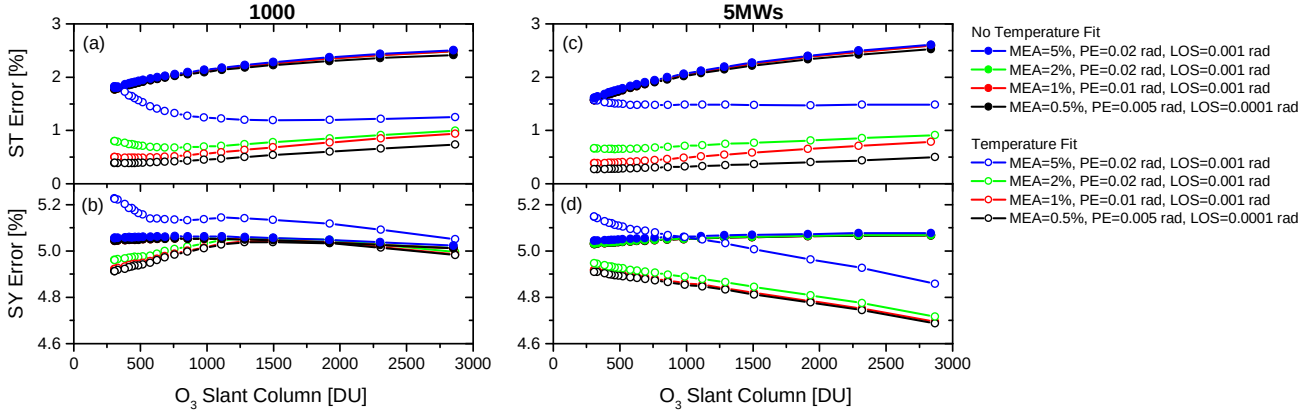



**Figure A1.** Estimated total statistical (ST) and systematic (SY) errors [%] for $O_3$ TCs retrieved from the set-ups 1000/1000T (a and b), and 5MWs/5MWsT (c and d) as a function of $O_3$ slant column [DU] for the FTIR measurements taken on $31^{st}$ August 2007 from SZAs between 84° (∼07:00 UT) and 21° (∼13:30 UT).

The effect of the different error sets is only noticeable when the temperature profile is simultaneously estimated with $O_3$ concentrations likely due to the interference between the ILS function and temperature retrieval. Under these conditions, total statistical error values range from ∼0.5% for uncertainties of 0.5% in the MEA and of 0.005 rad in the PE (representative of the well-aligned IFS 120/5HR instrument for the period 2008-2018) up to ∼1.5% for an MEA error of 5% and PE of 0.02 rad (representative of the unstable IFS 120M spectrometer). These estimated uncertainties reproduce well the changes observed in FTIR $O_3$ quality for the different periods when comparing to Brewer observations (recall Table 2). The cross-interference between the temperature fit and other error sources is also evident for the total systematic contributions, especially for the worst

https://doi.org/10.5194/amt-2021-67
Preprint. Discussion started: 3 May 2021


Atmospheric Measurement Techniques Discussions

scenario of ILS degradation (MEA uncertainty values of 5%). Note that the inconsistency for the 1000T set-up is also observed for the different ILS set errors, corroborating the findings discussed in Section 3.2.2.

*Data availability.* The FTIR and Brewer data are available by request from the corresponding authors, while the ECC ozone sondes are available from the NDACC archive (www.ndaccdemo.org/).

*Author contributions.* O.G. and E.S. designed and wrote the structure and methodology of the current paper, and computed the calculations required. M.S., F.H., and T.B. participated in the retrieval analysis. F.H. is the author of LINEFIT and PROFFIT codes. O.G., M.S. E.S., and E.S. taken the routine FTIR measurements, and performed the maintenance and quality-control of the FTIR instruments. A.R., S.F., and V.C. are responsible of maintenance and quality-control of the Brewer spectrometers, as well as of estimating the NDACC Brewer ozone observations. C.T. and N.P. are in charge of the ozone sonde programme at Izaña Observatory. Finally, all authors discussed the results and contributed to the final paper.

*Competing interests.* The authors declare no conflict of interest.

*Acknowledgements.* The research leading to these results has received funding from the Ministerio de Economía y Competitividad from Spain through the project INMENSE (CGL2016-80688-P) and by the Deutsche Forschungsgemeinschaft for the project MOTIV (Geschaftszeichen SCHN 1126/2-1).

https://doi.org/10.5194/amt-2021-67
Preprint. Discussion started: 3 May 2021


Atmospheric Measurement Techniques Discussions

## References

Austin, J. and Butchart, N.: Coupled chemistry–climate model simulations for the period 1980 to 2020: Ozone depletion and the start of ozone recovery, Quarterly Journal of the Royal Meteorological Society, 129, 3225–3249, https://doi.org/10.1256/qj.02.203, 2003.

Barret, B., de MazièRe, M., and Demoulin, P.: Retrieval and characterization of ozone profiles from solar infrared spectra at the Jungfraujoch, Journal of Geophysical Research, 107, 4788, https://doi.org/10.1029/2001JD001298, 2002.

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 Ozone Symposium, vol. 1, pp. 606–616, 1985.

Cuevas, E., Milford, C., Bustos, J. J., R., García, O. E., García, R. D., Gómez-Peláez, A. J., Guirado-Fuentes, C., Marrero, C., Prats, N., Ramos, R., Redondas, A., Reyes, E., Rivas-Soriano, P. P., Rodríguez, S., Romero-Campos, P. M., Torres, C. J., Schneider, M., Yela, M., Belmonte, J., del Campo-Hernández, R., Almansa, F., Barreto, A., López-Solano, C., Basart, S., Terradellas, E., Werner, E., Afonso, S., Bayo, C., Berjón, A., Carreño, V., Castro, N. J., Chinea, N., Cruz, A. M., Damas, M., De Ory-Ajamil, F., García, M., Gómez-Trueba, V., Hernández, C., Hernández, Y., Hernández-Cruz, B., León-Luís, S. F., López-Fernández, R., López-Solano, J., Parra, F., Rodríguez, E., Rodríguez-Valido, M., Sálamo, C., Sanromá, E., Santana, D., Santo Tomás, F., Sepúlveda, E., and Sosa, E.: Izaña Atmospheric Research Center Activity Report 2017-2018, Eds. Cuevas, E., Milford, C. and Tarasova, O.), State Meteorological Agency (AEMET), Madrid, Spain, and World Meteorological Organization (WMO), Geneva, Switzerland, WMO/GAW Report No. 247, 2019.

Cuevas, E., González, Y., Rodríguez, S., Guerra, J. C., Gómez-Peláez, A. J., Alonso-Pérez, S., Bustos, J., and Milford, C.: Assessment of atmospheric processes driving ozone variations in the subtropical North Atlantic free troposphere, Atmospheric Chemistry and Physics, 13, 1973–1998, https://doi.org/10.5194/acp-13-1973-2013, 2013.

Duflot, V., Baray, J.-L., Payen, G., Marquestaut, N., Posny, F., Metzger, J.-M., Langerock, B., Vigouroux, C., Hadji-Lazaro, J., Portafaix, T., De Mazière, M., Coheur, P.-F., Clerbaux, C., and Cammas, J.-P.: Tropospheric ozone profiles by DIAL at Maïdo Observatory (Reunion Island): system description, instrumental performance and result comparison with ozone external data set, Atmospheric Measurement Techniques, 10, 3359–3373, https://doi.org/10.5194/amt-10-3359-2017, 2017.

Eyring, V., Cionni, I., Bodeker, G. E., Charlton-Perez, A. J., Kinnison, D. E., Scinocca, J. F., Waugh, D. W., Akiyoshi, H., Bekki, S., Chipperfield, M. P., Dameris, M., Dhomse, S., Frith, S. M., Garny, H., Gettelman, A., Kubin, A., Langematz, U., Mancini, E., Marchand, M., Nakamura, T., Oman, L. D., Pawson, S., Pitari, G., Plummer, D. A., Rozanov, E., Shepherd, T. G., Shibata, K., Tian, W., Braesicke, P., Hardiman, S. C., Lamarque, J. F., Morgenstern, O., Pyle, J. A., Smale, D., and Yamashita, Y.: Multi-model assessment of stratospheric ozone return dates and ozone recovery in CCMVal-2 models, Atmospheric Chemistry and Physics, 10, 9451–9472, https://doi.org/10.5194/acp-10-9451-2010, 2010.

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., Sepúlveda, E., Schneider, M., Hase, F., August, T., Blumenstock, T., Kühl, S., Munro, R., Gómez-Peláez, A. J., Hultberg, T., Redondas, A., Barthlott, S., Wiegele, A., González, Y., and Sanromá, E.: Consistency and quality assessment of the Metop-A/IASI and Metop-B/IASI operational trace gas products ($O_3$, CO, $N_2O$,

https://doi.org/10.5194/amt-2021-67
Preprint. Discussion started: 3 May 2021




$CH_4$, and $CO_2$) in the subtropical North Atlantic, Atmospheric Measurement Techniques, 9, 2315–2333, https://doi.org/10.5194/amt-9-2315-2016, 2016.

Gratien, A., Picquet-Varrault, B., Orphal, J., Doussin, J.-F., and Flaud, J.-M.: New Laboratory Intercomparison of the Ozone Absorption Coefficients in the Mid-infrared (10 $\mu$m) and Ultraviolet (300–350 nm) Spectral Regions, Journal of Physical Chemistry, 114, 10 045–10 048, https://doi.org/10.1021/jp103992f, 2010.

Gröbner, J., Redondas, A., Weber, M., and Bais, A.: Final report of the project Traceability for atmospheric total column ozone (ENV59, ATMOZ), Tech. rep., EURAMET, https://www.euramet.org/research-innovation/search-research-projects/details/project/traceability-for-atmospheric-total-column-ozone/, 2017.

Guinet, M., Mondelain, D., Janssen, C., and Camy-Peyret, C.: Laser spectroscopic study of ozone in the 100←000 band for the SWIFT instrument, Journal of Quantitative Spectroscopy and Radiative Transfer, 111, 961–972, https://doi.org/10.1016/j.jqsrt.2010.01.011, 2010.

Hase, F.: Error Estimation in PROFFIT revisited, in: NDSC IRWG Meeting, Tenerife, Spain, 2007.

Hase, F.: Improved instrumental line shape monitoring for the ground-based, high-resolution FTIR spectrometers of the Network for the Detection of Atmospheric Composition Change, Atmospheric Measurement Techniques, 5, 603–610, https://doi.org/10.5194/amt-5-603-2012, 2012.

Hase, F., Hannigan, J., Coffey, M., Goldman, A., Hopfner, M., Jones, N., Rinsland, C., and Wood, S.: Intercomparison of retrieval codes used for the analysis of high-resolution, ground-based FTIR measurements, Journal of Quantitative Spectroscopy and Radiative Transfer, 87, 25–52, https://doi.org/10.1016/j.jqsrt.2003.12.008, 2004.

IRWG: Uniform Retrieval Parameter Summary, http://www.acom.ucar.edu/irwg/IRWG_Uniform_RP_Summary-3.pdf, 2014.

Komhyr, W.: Operations handbook - Ozone measurements to 40 km altitude with model 4AECC-ozone sondes, Tech. rep., National Oceanic and Atmospheric Administration, Technical Memorandum ERL-ARL-149, 1986.

Lanzante, J. R.: Resistant, robust and non-parametric techniques or the analysis of climate data: theory and examples, including applications to historical radiosonde station data, International Journal of Climatology, 16, 1197–1226, 1996.

León-Luis, S. F., Redondas, A., Carreño, V., López-Solano, J., Berjón, A., Hernández-Cruz, B., and Santana-Díaz, D.: Internal consistency of the Regional Brewer Calibration Centre for Europe triad during the period 2005–2016, Atmospheric Measurement Techniques, 11, 4059–4072, https://doi.org/10.5194/amt-11-4059-2018, 2018.

Lindenmaier, R., Batchelor, R. L., Strong, K., Fast, H., Goutail, F., Kolonjari, F., C. T. McElroy, R. L. M., and Walker, K. A.: An evaluation of infrared microwindows for ozone retrievals using the Eureka Bruker 125HR Fourier transform spectrometer, Journal of Quantitative Spectroscopy and Radiative Transfer, 111, 569–585, 2010.

Piquet-Varrault, B., Orphal, J., Doussin, J. F., Carlier, P., and Flaud, J. M.: Laboratory Intercomparaison of the Ozone Absorption Coefficients in the Mid-infrared (10$\mu$m) and Ultarviolet (300– 350 nm) Spectral Regions, Journal of Physical Chemistry, 109, 1008–1014, https://doi.org/10.1021/jp0405411, 2005.

Redondas, A. and Cede, A.: A.: Brewer algorithm sensitivity analysis, in: SAUNA workshop, Puerto de la Cruz, Tenerife, Vol. 286, 2006.

Redondas, A., Evans, R., Stuebi, R., Köhler, U., and Weber, M.: Evaluation of the use of five laboratory-determined ozone absorption cross sections in Brewer and Dobson retrieval algorithms, Atmospheric Chemistry and Physics, 14, 1635–1648, https://doi.org/10.5194/acp-14-1635-2014, 2014.

Rodgers, C.: Inverse Methods for Atmospheric Sounding: Theory and Praxis, World Scientific Publishing Co., Singapore, 2000.

Rothman, L., Gordon, I., Barbe, A., Benner, D., Bernath, P., Birk, M., Boudon, V., Brown, L., Campargue, A., Champion, J.-P., Chance, K., Coudert, L., Dana, V., Devi, V., Fally, S., Flaud, J.-M., Gamache, R., Goldman, A., Jacquemart, D., Kleiner, I., Lacome, N., Lafferty, W.,

https://doi.org/10.5194/amt-2021-67
Preprint. Discussion started: 3 May 2021




Mandin, J.-Y., Massie, S., Mikhailenko, S., Miller, C., Moazzen-Ahmadi, N., Naumenko, O., Nikitin, A., Orphal, J., Perevalov, V., Perrin, A., Predoi-Cross, A., Rinsland, C., Rotger, M., Šimečková, M., Smith, M., Sung, K., Tashkun, S., Tennyson, J., Toth, R., Vandaele, A., and Auwera, J. V.: The HITRAN 2008 molecular spectroscopic database, Journal of Quantitative Spectroscopy and Radiative Transfer, 110, 533 – 572, https://doi.org/http://dx.doi.org/10.1016/j.jqsrt.2009.02.013, 2009.

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 and Physics, 8, 63–71, https://doi.org/10.5194/acp-8-63-2008, 2008.

Schneider, M., Blumenstock, T., Chipperfield, M. P., Hase, F., Kouker, W., Reddmann, T., Ruhnke, R., Cuevas, E., and Fischer, H.: Subtropical trace gas profiles determined by ground-based FTIR spectroscopy at Izaña (28ºN, 16ºW): Five-year record, error analysis, and comparison with 3-D CTMs, Atmospheric Chemistry and Physics, 5, 153–167, https://doi.org/10.5194/acp-5-153-2005, 2005.

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 and Physics, 8, 5535–5550, https://doi.org/10.5194/acp-8-5535-2008, 2008a.

Schneider, M., Hase, F., Blumenstock, T., Redondas, A., and Cuevas, E.: Quality assessment of $O_3$ profiles measured by a state-of-the-art ground-based FTIR observing system, Atmospheric Chemistry and Physics, 8, 5579–5588, 2008b.

Schneider, M., Barthlott, S., Hase, F., González, Y., Yoshimura, K., García, O. E., Sepúlveda, E., Gomez-Pelaez, A., Gisi, M., Kohlhepp, R., Dohe, S., Blumenstock, T., Wiegele, A., Christner, E., Strong, K., Weaver, D., Palm, M., Deutscher, N. M., Warneke, T., Notholt, J., Lejeune, B., Demoulin, P., Jones, N., Griffith, D. W. T., Smale, D., and Robinson, J.: Ground-based remote sensing of tropospheric water vapour isotopologues within the project MUSICA, Atmospheric Measurement Techniques, 5, 3007–3027, https://doi.org/10.5194/amt-5-3007-2012, 2012.

Viatte, C., Schneider, M., Redondas, A., Hase, F., Eremenko, M., Chelin, P., Flaud, J.-M., Blumenstock, T., and Orphal, J.: Comparison of ground-based FTIR and Brewer $O_3$ total column with data from two different IASI algorithms and from OMI and GOME-2 satellite instruments, Atmospheric Measurement Techniques, 4, 535–546, https://doi.org/10.5194/amt-4-535-2011, 2011.

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 and 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 and Physics, 15, 2915–2933, https://doi.org/10.5194/acp-15-2915-2015, 2015.

Weatherhead, E. C., Reinsel, G. C., Tiao, G. C., Jackman, C. H., Bishop, L., Frith, S. M. H., DeLuisi, J., Keller, T., Oltmans, S. J., Fleming, E. L., Wuebbles, D. J., Kerr, J. B., Miller, A. J., Herman, J., McPeters, R., Nagatani, R. M., and Frederick, J. E.: Detecting the recovery of total column ozone, Journal of Geophysical Research, 105, 22 201–22 210, https://doi.org/10.1029/2000JD900063, 2000.

WMO: Scientific Assessment of Ozone Depletion: 2014, Global Ozone Research and Monitoring Project–Report No. 55, Tech. rep., World Meteorological Organization, Geneva, Switzerland, 2014a.

WMO: Quality assurance and quality control for ozonesonde measurements in GAW, World Meteorological Organization (WMO)-Report No. 201, Tech. rep., Smit, H.G.J., and ASOPOS panel (eds.), World Meteorological Organization, Geneva, Switzerland, 2014b.

WMO: Scientific Assessment of Ozone Depletion: 2018, Global Ozone Research and Monitoring Project–Report No. 58, Tech. rep., World Meteorological Organization, Geneva, Switzerland, 2018.