



# Improved ozone monitoring by ground-based FTIR spectrometry

Omaira Elena García<sup>1</sup>, Esther Sanromá<sup>1,a</sup>, Matthias Schneider<sup>2</sup>, Frank Hase<sup>2</sup>, Sergio Fabián León-Luis<sup>1,b</sup>, Thomas Blumenstock<sup>2</sup>, Eliezer Sepúlveda<sup>1</sup>, Alberto Redondas<sup>1</sup>, Virgilio Carreño<sup>1</sup>, Carlos Torres<sup>1</sup>, and Natalia Prats<sup>1</sup>

<sup>1</sup>Izaña Atmospheric Research Centre (IARC), State Meteorological Agency of Spain (AEMet), Santa Cruz de Tenerife, Spain

<sup>2</sup>Institute of Meteorology and Climate Research (IMK-ASF), Karlsruhe Institute of Technology (KIT), Karlsruhe, Germany

<sup>a</sup>now at: Employment Observatory of the Canary Islands (OBECAN), Santa Cruz de Tenerife, Spain

<sup>b</sup>now at: TRAGSATEC, Madrid, Spain

**Correspondence:** Omaira Elena García (ogarcia@amet.es)

Received: 5 March 2021 – Discussion started: 3 May 2021

Revised: 1 April 2022 – Accepted: 4 April 2022 – Published: 28 April 2022

**Abstract.** Accurate observations of atmospheric ozone ( $O_3$ ) are essential to monitor in detail its key role in 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 effects of two of the most influential factors have been investigated: the inclusion of a simultaneous atmospheric temperature profile fit and the spectral  $O_3$  absorption lines used for the retrievals (the broad spectral region of  $1000\text{--}1005\text{ cm}^{-1}$  and single micro-windows between  $991$  and  $1014\text{ cm}^{-1}$ ). Additionally, the water vapour ( $H_2O$ ) interference in  $O_3$  retrievals has been evaluated, with the aim of providing an improved  $O_3$  strategy that minimises its impact and, therefore, could be applied at any NDACC FTIR station under different humidity conditions. 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 a simultaneous temperature retrieval with the optimal selection of single  $O_3$  micro-windows results in superior FTIR  $O_3$  products, with a precision of better than  $0.6\%$ – $0.7\%$  for  $O_3$  TCs as compared to coincident NDACC Brewer observations taken as a reference. However, this improvement can only be achieved provided the FTIR spectrometer is properly characterised and stable over time. For unstable instruments, the temperature fit is found to exhibit a strong negative influence on  $O_3$  retrievals due to the increase in the cross-interference between the temperature re-

trieval and instrumental performance (given by the instrumental line shape function and measurement noise), which leads to a worsening of the precision of FTIR  $O_3$  TCs of up to  $2\%$ . 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 the  $O_3$  vertical distribution, especially over long-term timescales.

## 1 Introduction

Monitoring 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 atmospheric chemistry. 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; it affects the oxidising capacity of the atmosphere; and it is a phytotoxicant that is harmful to public health (Cuevas et al., 2013; WMO, 2014a, 2018; Gaudel et al., 2018, and references therein).

Stratospheric O<sub>3</sub> abundances have shown a significant decrease in global levels from the 1980s to the 1990s, mainly attributable to the increase in anthropogenic emissions 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<sub>3</sub> decay by controlling ODS emissions: O<sub>3</sub> concentrations have approximately stabilised since stratospheric ODS abundances reached their maximum at the end of the 1990s. As a result, the global O<sub>3</sub> 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<sub>3</sub> concentrations are affected not only by the presence of ODSs, but also by a wide variety of factors (increases in greenhouse gas concentrations, changes in the Brewer–Dobson circulation and stratospheric temperatures, etc.), making it very challenging to predict how, when, and where the O<sub>3</sub> recovery will take place. In the troposphere, since O<sub>3</sub> is highly variable (depending on time period, region, and proximity to fresh O<sub>3</sub> precursor emissions), there is no consistent picture of O<sub>3</sub> tropospheric changes around the world (Steinbrecht et al., 2017; Gaudel et al., 2018; WMO, 2018). Hence, high-quality and long-term O<sub>3</sub> measurements are essential for further improving our understanding of the O<sub>3</sub> response to natural and anthropogenic forcings, as well as to estimate consistent trends at a global scale (Vigouroux et al., 2015; Gaudel et al., 2018; GCOS, 2021).

Within NDACC (Network for the Detection of Atmospheric Composition Change, <http://www.ndaccdemo.org>, last access: 31 March 2022), 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 currently capable of providing both high-quality O<sub>3</sub> total column (TC) amounts and low-resolution O<sub>3</sub> vertical volume mixing ratio (VMR) profiles at about 20 sites (e.g. Barret et al., 2002; Schneider and Hase, 2008; Schneider et al., 2008a, b; Vigouroux et al., 2008; Viatte et al., 2011; García et al., 2012, 2014; Vigouroux et al., 2015, and references therein). In recent years, the NDACC Infrared Working Group (IRWG, <http://www2.acom.ucar.edu/irwg>, last access: 31 March 2022) has made considerable efforts to standardise the data acquisition protocols and inversion strategies used to derive O<sub>3</sub> concentrations at the different NDACC stations and, hence, produce uniform and consistent O<sub>3</sub> datasets (Hase et al., 2004; IRWG, 2014; Vigouroux et al., 2015). Nonetheless, scientific discussions seeking improvements in O<sub>3</sub> monitoring and network-wide consistency are ongoing.

In this context, the present paper examines the effect of using different retrieval approaches on the quality of the FTIR O<sub>3</sub> products, with the aim of providing an improved O<sub>3</sub> strategy that could be applied at any NDACC FTIR station. The influences of two of the most important settings are assessed:

the spectral region used for O<sub>3</sub> retrievals and the simultaneous fitting of the atmospheric temperature profile. To our knowledge, so far, such analysis has been approached separately in most of the studies present in the literature, or has not been addressed in detail yet. Previous studies have shown, for example, that optimised selection of the O<sub>3</sub> absorption lines or the inclusion of an additional temperature fitting significantly improves the precision of FTIR O<sub>3</sub> TCs 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 those works.

The analysis was performed at the O<sub>3</sub> super-site of Izaña Observatory (IZO), where ground-based FTIR observations have been carried out coincidentally with other high-quality O<sub>3</sub> measurement techniques since 1999. By using those data, a comprehensive assessment of the precision and long-term consistency of new O<sub>3</sub> retrieval strategies from ground-based FTIR spectrometry can be carried out. To this end, the current paper is structured as follows. Section 2 describes the Izaña Observatory, FTIR measurements, and ancillary data considered to assess the quality of the new FTIR O<sub>3</sub> products (Brewer TC observations and electrochemical concentration cell (ECC) vertical sondes). Section 3 presents the different FTIR O<sub>3</sub> 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<sub>3</sub> TCs and VMR profiles by comparing them to the independent O<sub>3</sub> datasets. Finally, Sect. 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 (28.3° N, 16.5° W) at an altitude of 2370 m a.s.l. The observatory is managed by the Izaña Atmospheric Research Centre (IARC, <https://izana.aemet.es>, last access: 31 March 2022), which belongs to the State Meteorological Agency of Spain (AEMet, <https://www.aemet.es>, last access: 31 March 2022). 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., 2020, and references therein).

For many years, IZO has run a comprehensive O<sub>3</sub> monitoring programme by using different measurement techniques: FTIR, Brewer, and DOAS (differential optical absorption spectroscopy) spectrometers, as well as ECC O<sub>3</sub> sondes and in situ ultraviolet photometric analysers. The first four techniques routinely contribute to NDACC, with the aim

being to monitoring changes in the troposphere and stratosphere with an emphasis on the long-term evolution of the O<sub>3</sub> layer, while the in situ O<sub>3</sub> records are taken in the framework of the WMO/GAW (World Meteorological Organization/Global Atmospheric Watch) programme. Please refer to Cuevas et al. (2020) 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 due to 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, <https://www.kit.edu>, last access: 31 March 2022) and AEMet-IARC. In 2005, this instrument was replaced with an 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, encompassing the operation of the two FTIR instruments, have been used.

Among the activities of NDACC, the IZO FTIR spectrometer records direct solar absorption spectra in the middle-infrared spectral region, i.e. 740–4250 cm<sup>-1</sup> (corresponding to 13.5–2.4 μm), by using a set of different field stops, narrow-bandpass filters, and detectors. Nevertheless, for O<sub>3</sub> retrievals, only the 960–1015 cm<sup>-1</sup> spectral region is considered, which is measured with 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<sup>-1</sup> (250 cm maximum optical path difference, OPD, OPD<sub>max</sub>) until April 2000, and at 0.005 cm<sup>-1</sup> (OPD<sub>max</sub> = 180 cm) subsequently. 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 and thereby the acquisition of one spectrum takes about 10 min.

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's location, 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<sub>3</sub> 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 García et al. (2021).

In order to characterise the instrumental performance of the IZO FTIR spectrometers, the instrumental line shape (ILS) function has been routinely monitored about every 2 months since 1999 using low-pressure N<sub>2</sub>O-cell measurements and LINEFIT software (v14.5), as detailed in Hase

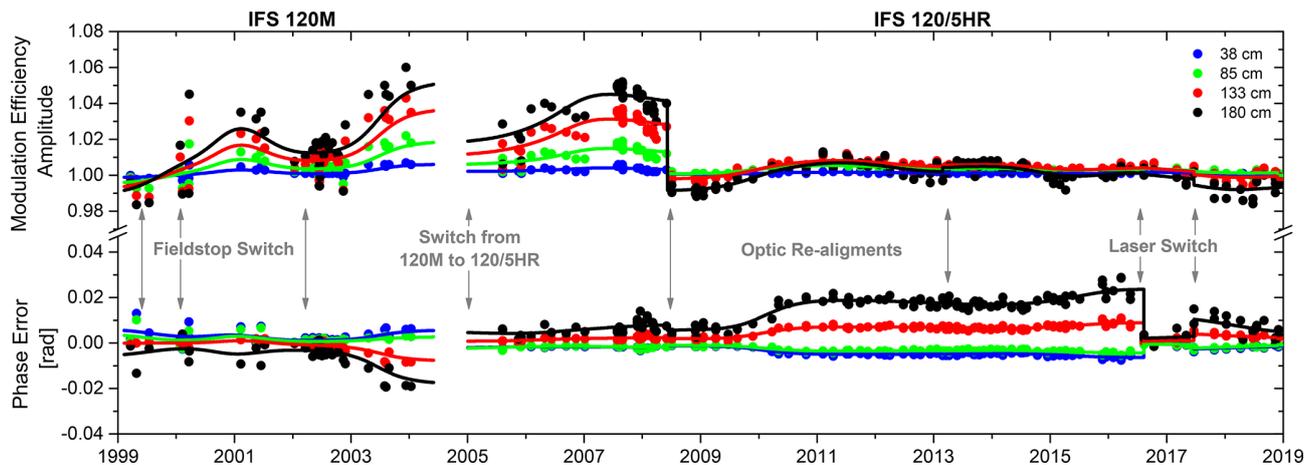
(2012). This ILS treatment ensures the independence of the FTIR trace gas retrievals and instrumental characterisation, but it also allows the instrumental alignment and its temporal stability to be verified. Figure 1 depicts the time series of the ILS's modulation efficiency amplitude (MEA) and phase error (PE) parameters between 1999 and 2018, as used for O<sub>3</sub> retrievals. Three periods with different features affecting the IZO FTIR measurements can be distinguished: (1) during 1999–2004, although N<sub>2</sub>O-cell measurements were routinely carried out, the ILS estimation is imprecise due to the instability of the IFS 120M spectrometer; (2) during 2005–May 2008, the IFS 120/5HR instrument exhibits a gradual temporal drift, but the ILS function is properly assessed; and (3) during June 2008–2018, the IFS 120/5HR instrument is optically well-aligned (the ILS is nearly nominal). Therefore, these three periods will be independently analysed in the present work in order to examine the influence of instrument status on FTIR O<sub>3</sub> 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, 2 years later, the RBCC-E (Regional Brewer Calibration Centre for Europe, <http://www.rbcc-e.org>, last access: 31 March 2022) of the WMO/GAW programme was established at IZO. By recording direct solar absorption spectra in the ultraviolet spectral region, IZO RBCC-E reference instruments can provide O<sub>3</sub> TCs with a total uncertainty (standard uncertainty,  $k = 1$ ) of 1.2%–1.5% (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).

The O<sub>3</sub> sonde programme on Tenerife, also run by AEMet, started in November 1992, and has operated since March 2001 within the framework of NDACC. The O<sub>3</sub> sounding is based on the ECC, which senses O<sub>3</sub> as it reacts with a dilute solution of potassium iodide (KI) to produce an electrical current proportional to the atmospheric O<sub>3</sub> concentrations (Komhyr, 1986). 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 the Botanic Observatory (13 km north of IZO, 114 m a.s.l.). The expected total uncertainty of the ECC sondes is ±5%–15% in the troposphere and ±5% in the stratosphere (WMO, 2014b), which is a composite of different instrumental error contributions (i.e. sensor and background current, conversion efficiency, etc.).

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



**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<sub>3</sub> measurement settings (filter 6 and MCT detector) of the IZO FTIR spectrometers between 1999 and 2018. Data points represent individual N<sub>2</sub>O-cell measurements and solid lines depict smoothed MEA and PE curves. Solid grey arrows indicate punctual interventions for the instruments: different changes of field stops between 1999 and 2004, the 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.

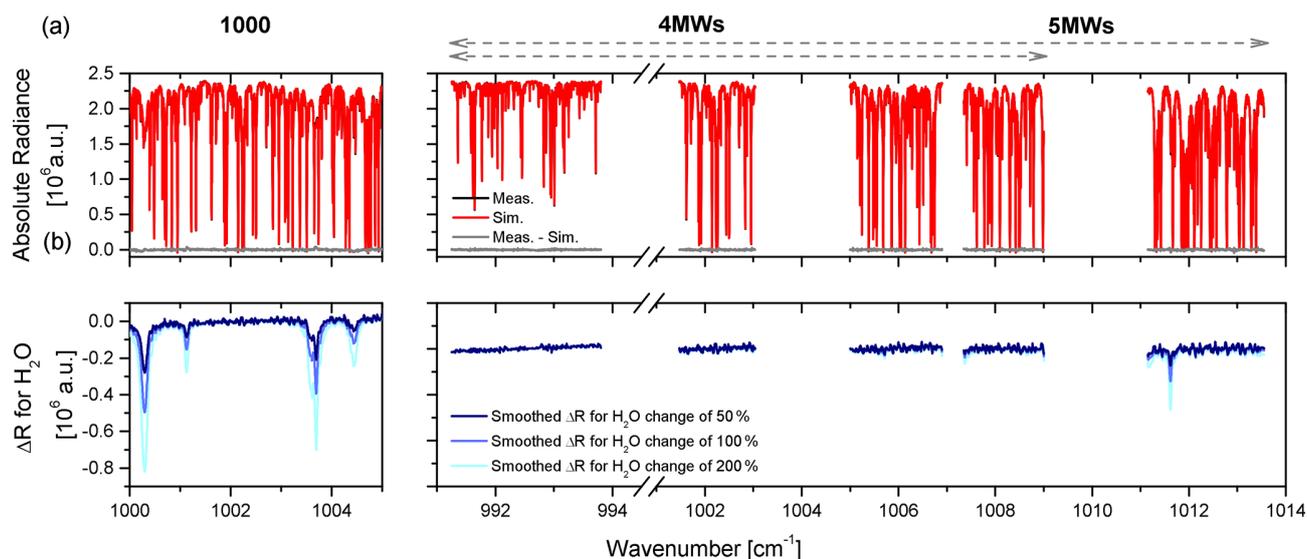
### 3 FTIR ozone observations

#### 3.1 Ozone retrieval strategies

To analyse the influence of the spectral region and the simultaneous temperature fit on the quality of FTIR O<sub>3</sub> products, six different approaches have been defined. They combine three spectral regions and the possibility of performing a simultaneous temperature retrieval (referred to as retrieval setups 1000, 4MWs, 5MWs, 1000T, 4MWsT, and 5MWsT hereafter; Fig. 2). Setup 1000 uses a broad spectral window covering 1000–1005 cm<sup>-1</sup>, which is the one recommended by the NDACC IRWG (IRWG, 2014). This spectral region has been traditionally used by the FTIR community when reporting high-quality O<sub>3</sub> products (e.g. Barret et al., 2002; Schneider et al., 2008a; Vigouroux et al., 2008; Lindenmaier et al., 2010; García et al., 2012; Vigouroux et al., 2015). Setup 5MWs uses five single micro-windows between 991 and 1014 cm<sup>-1</sup>, 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<sub>3</sub> estimations than those retrieved from the broad 1000–1005 micro-window when compared to independent measurements. Setup 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<sub>3</sub> absorption lines contained in this region, especially at high O<sub>3</sub> concentrations and low solar elevations. Setups 1000T, 4MWsT, and 5MWsT use the same micro-windows as setups 1000, 4MWs, and 5MWs respectively, but an optimal estimation of the atmospheric temperature profile is simultaneously carried out. To this end, four CO<sub>2</sub> micro-windows are added between 962.80 and 969.60 cm<sup>-1</sup> according to García et al. (2012).

With the exception of the spectral region and temperature treatment, the retrieval strategy is identical for the six approaches. The O<sub>3</sub> VMR profiles are derived from the measured solar absorption spectra by means of the PROFFIT code (PROFile FIT, Hase et al., 2004), using an ad hoc Tikhonov–Phillips slope constraint (TP1 constraint) on a logarithmic scale. Since O<sub>3</sub> concentrations are very variable around the tropopause, logarithmic inversion has proved to be superior to the linear approach (e.g. Schneider and Hase, 2008; Schneider et al., 2008a, b). Then, the O<sub>3</sub> 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 the NDACC IRWG recommendations (IRWG, 2014):

- The interfering species considered are H<sub>2</sub>O, CO<sub>2</sub>, C<sub>2</sub>H<sub>4</sub>, and the main O<sub>3</sub> isotopologues (666, 686, 668, 667, and 676 in HITRAN notation). Spectroscopic line parameters of these absorbers are taken from HITRAN 2008 (Rothman et al., 2009), with a 2009 update for H<sub>2</sub>O (<http://www.cfa.harvard.edu>, last access: 31 March 2022).
- All setups 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>, last access: 31 March 2022) version 6, generated by the NCAR (National Center for Atmospheric Research, James Hannigan, personal communication, 2014).
- All setups apply the actual ILS time series evaluated from independent N<sub>2</sub>O-cell measurements (Fig. 1).



**Figure 2.** (a) Spectral regions considered in the different FTIR  $\text{O}_3$  retrieval strategies: the broad window used in the setups 1000 and 1000T (encompassing the  $1000\text{--}1005\text{ cm}^{-1}$  spectral region) and the four and five micro-windows used in the setups 4MWs/4MWsT (between  $\sim 991$  and  $1009\text{ cm}^{-1}$ ) and 5MWs/5MWsT (between  $\sim 991$  and  $1014\text{ cm}^{-1}$ ), respectively. Red and black lines depict, respectively, examples of the simulated and measured spectra taken on 31 August 2007 (at a solar zenith angle (SZA) of  $\approx 50^\circ$ , an  $\text{O}_3$  amount in the slant column ( $\text{O}_3$  SC) of 390 DU, and an  $\text{H}_2\text{O}$  TC of 10.7 mm). Grey lines show the difference between the measurement and simulation. (b) Spectral changes in the measured FTIR radiances ( $\Delta R$ ) due to changes in the  $\text{H}_2\text{O}$  content of 50 % (TC of 16.1 mm), 100 % (TC of 21.5 mm), and 200 % (TC of 32.3 mm).

- The pressure and temperature profiles for forward simulations are taken from the NCEP (National Centers for Environmental Predictions) 12:00 UT daily database.
- For those approaches performing a simultaneous optimal estimation of the atmospheric temperature profile, the NCEP 12:00 UT daily temperature profiles are used as the a priori profiles. The a priori temperature covariance matrix ( $\mathbf{S}_{aT}$ ) is constructed following Schneider et al. (2008a).

Given the importance of  $\text{H}_2\text{O}$  absorption across the infrared spectrum, the treatment of  $\text{H}_2\text{O}$  in  $\text{O}_3$  retrievals should be carefully considered in the inversion strategy, as illustrated in Fig. 2. This figure shows an example of the changes in the FTIR radiances for the  $\text{O}_3$  spectral micro-windows due to changes in the  $\text{H}_2\text{O}$  content of 50 %, 100 %, and 200 %. These values, which correspond to extreme conditions at IZO, might account for typical  $\text{H}_2\text{O}$  contents and variations at sites with greater humidity. As observed, the spectral signatures of  $\text{H}_2\text{O}$  variations are much stronger in the broad 1000 spectral region than in the narrow micro-windows (4MWs/5MWs), indicating that the quality of the  $\text{O}_3$  products in that region strongly depends on the correct interpretation of the spectroscopic  $\text{H}_2\text{O}$  signal. Therefore, in order to minimise interference errors due to  $\text{H}_2\text{O}$ , a two-step inversion strategy has been applied (García et al., 2012, 2014): firstly, the actual  $\text{H}_2\text{O}$  profile is derived using a dedicated  $\text{H}_2\text{O}$  retrieval (Schneider et al., 2012), and then the

$\text{O}_3$  retrieval is simultaneously performed with an  $\text{H}_2\text{O}$  scaling fit, which uses the previously derived  $\text{H}_2\text{O}$  state. The remaining interfering gases are simultaneously estimated with  $\text{O}_3$  in the second step. As discussed in detail in Appendix A, the  $\text{H}_2\text{O}$  cross-interference is reduced by the two-step strategy when the temperature retrieval is considered, which suggests that this approach could be valid for humid FTIR sites.

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 would likely be introduced, for example, by thin clouds) (García et al., 2016). These two quality flags are applied independently on the six  $\text{O}_3$  datasets, and only those measurements available for all setups are considered in subsequent analysis. This leads to a total of 5393  $\text{O}_3$  observations between 1999 and 2018 that 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 the retrieved and true states must be provided together with the retrieved vertical profile. This information is theoretically characterised by the averaging kernel matrix ( $\mathbf{A}$ ) ob-

tained 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 called the degrees of freedom for signal, DOFS) gives the number of independent O<sub>3</sub> layers detectable by the remote-sensing FTIR instrument. As an example, Fig. 3a depicts the **A** rows for the 5MWs setup for the measurement of Fig. 2, while Table 1 summarises the DOFS statistics for the six retrieval strategies considered.

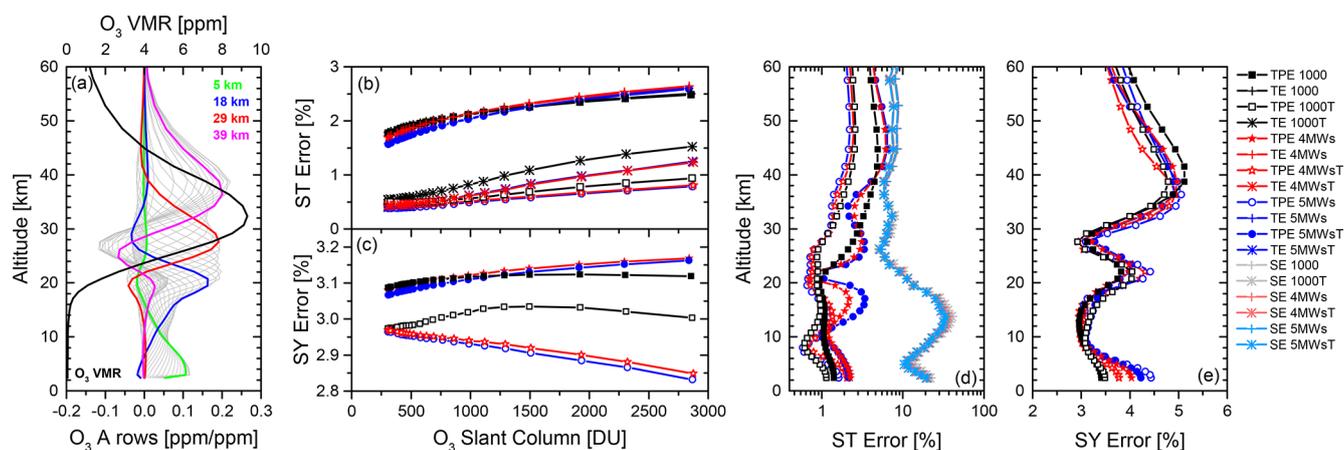
The **A** rows are quite similar for all setups, with a median total DOFS value of  $\sim 4$ , meaning that the FTIR system is able to roughly resolve four independent atmospheric O<sub>3</sub> layers: the troposphere (2.37–13 km), the upper troposphere/lower stratosphere (UTLS) or tropopause region (12–23 km), the middle stratosphere around the ozone maximum (22–29 km), and the upper stratosphere (28–42 km). However, the total DOFS values are found to be greater for those setups using narrower micro-windows than for the broad spectral window (see Table 1), whereby the former configurations seem to offer better vertical sensitivity (especially the 5MWs setup). This pattern is independent of 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 between the instruments also reveals, as expected, a lower sensitivity to the O<sub>3</sub> concentrations of the IFS 120M spectra as compared to the IFS 120/5HR measurements. Overall, the total DOFS values 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 the measured spectra is then split into O<sub>3</sub> and temperature retrievals (the retrieved state vector space is not perfectly orthogonal). Likewise, the differences between both instruments 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 the DOFS analysis, the fitting residuals are smaller for those setups that use narrow micro-windows and apply the temperature fit, allowing a more detailed interpretation of the 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 the IFS 120/5HR data. However, it is fair to admit that the differences among retrieval strategies lie within the respective error confidence intervals, so 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 present in all setups (1001.47–1003.04 cm<sup>-1</sup>).

### 3.2.2 Uncertainty analysis

The characterisation of the different FTIR O<sub>3</sub> products is completed by performing an uncertainty analysis, which

evaluates how different error sources could be propagated into the retrieved products. The theoretical error assessment carried out in the present paper is based on Rodgers (2000) and analytically performed by the PROFFIT package. The Rodgers formalism distinguishes three types of error: (1) smoothing error (SE) 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. Given that SE can be considered an inherent characteristic of the remote-sensing technique, it is not included in the uncertainty assessment suggested by the NDACC IRWG (IRWG, 2014). Therefore, it has been separately considered in this work by distinguishing the total parameter error (TPE), which is calculated as the square root of the quadratic sum of all error sources considered with the exception of the SE, and the total error (TE), which considers both the TPE and SE. A detailed description of the uncertainty assessment is given in Appendix B.

In order to assess the effect of O<sub>3</sub> absorption signatures on the uncertainty budget, the dependence of the estimated errors on the O<sub>3</sub> slant column (SC) amounts for each retrieval setup has been examined. This analysis allows possible inconsistencies between the setups or saturations of O<sub>3</sub> absorption lines at high O<sub>3</sub> concentrations to be detected. As shown in Fig. 3, both statistical and systematic uncertainties do depend on the O<sub>3</sub> spectroscopic signatures for all setups due to the increase in most of the error sources considered at larger O<sub>3</sub> SCs (see details in Appendix B). Figure 3 also documents that the inclusion of a simultaneous temperature retrieval significantly improves the theoretical performance for all FTIR O<sub>3</sub> products. Although this fit generates a negative cross-interference with the ILS, measurement noise, and smoothing error (especially for the 1000T setup, Fig. B1), in return, the temperature error contribution is nearly eliminated, decreasing the total ST budget by  $\sim 1$  %. The TPE and TE range from 1.5 % to 2.5 % (between 250 and 3000 DU) for setups without a simultaneous temperature fit, while the TPE varies from 0.5 % to 1.0 % (or to 1.5 % for the TE due to the influence of the SE) when including the temperature retrieval. The total systematic contributions also drop by 0.3 % (at high O<sub>3</sub> SCs) with mean values of  $\sim 3$  %. It is worth highlighting that an inconsistency between the 4MWsT/5MWsT and 1000T setups has been detected in the systematic uncertainty budget, which is determined by the spectroscopy errors. For the 1000T configuration, the spectroscopic SY error exhibits a reverse smile curve with O<sub>3</sub> concentrations, and is considerably greater than for the narrow micro-window setups. This result might point to a possible saturation of the deeper O<sub>3</sub> lines contained in the broad window or some inconsistency in the spectroscopy parameters. For example, an erroneous parameterisation of the temperature dependence of the O<sub>3</sub> line width may produce systematic differences between actual and retrieved temperature profiles (Schneider



**Figure 3.** Summary of the theoretical quality assessment. **(a)** Example of averaging kernel (**A**) rows for the 5MWs setup on a logarithmic scale for the measured spectrum of Fig. 2. Coloured lines highlight the **A** rows at altitudes of 5, 18, 29, and 39 km, which are representative of the four layers detectable by the FTIR instrument (total DOFS of 4.41). The retrieved O<sub>3</sub> VMR profile is also shown. **(b)** Statistical (ST) and **(c)** systematic (SY) contributions of the total parameter error (TPE, in %) and total error (TE, in %) for O<sub>3</sub> TCs retrieved from the setups 1000/1000T, 4MWs/4MWsT, and 5MWs/5MWsT as a function of O<sub>3</sub> slant column (DU) for measurements taken on 31 August 2007 from SZAs between 84° (~07:00 UT) and 21° (~13:30 UT). TPE is computed as the square root of the quadratic sum of all ST and SY error sources considered with the exception of the smoothing error (SE), while the TE considers the TPE and SE. Examples of the ST **(d)** and SY **(e)** contributions of the TPE and SE profiles (%) for all setups for the measured spectrum of Fig. 2.

**Table 1.** Summary of statistics of the DOFS and fitting residuals for the setups 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 the median (*M*) and standard deviation (*σ*) 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. The strategies showing the best performance, in terms of largest DOFS and smallest residuals, are highlighted in bold for each period.

Setup	DOFS				Residuals ( $\times 10^{-3}$ )			
	1999–2004 <i>M</i> , <i>σ</i>	2005–2008 <i>M</i> , <i>σ</i>	2008–2018 <i>M</i> , <i>σ</i>	1999–2018 <i>M</i> , <i>σ</i>	1999–2004 <i>M</i> , <i>σ</i>	2005–2008 <i>M</i> , <i>σ</i>	2008–2018 <i>M</i> , <i>σ</i>	1999–2018 <i>M</i> , <i>σ</i>
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	<b>4.29</b> , 0.28	<b>4.56</b> , 0.15	<b>4.52</b> , 0.12	<b>4.51</b> , 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	<b>2.55</b> , 0.86	<b>2.68</b> , 0.54	<b>2.70</b> , 0.91
5MWsT	4.10, 0.33	4.42, 0.20	4.35, 0.15	4.35, 0.21	<b>3.45</b> , 1.90	2.56, 0.87	<b>2.68</b> , 0.54	<b>2.70</b> , 0.91

and Hase, 2008), therefore affecting the absolute value of O<sub>3</sub> FTIR products.

Vertically, the most important contribution is the SE reaching ~40% in the UTLS region (Fig. 3d), where the O<sub>3</sub> concentrations are very variable and the profile might be highly structured. The FTIR system is not able to resolve such fine vertical structures. Excluding the SE, the statistical TPE profiles are strongly linked to the atmospheric temperature, with maximal errors beyond the UTLS region (where the maximum FTIR sensitivity and the largest O<sub>3</sub> concentrations are also located, see Fig. 3a). This pattern is consistently observed for both setups with and without fitting the atmospheric temperature profiles. However, the error values drastically drop when considering the temperature in the retrieval

procedure. TPE values of 1.0%–1.5% are expected for the 1000T, 4MWsT, and 5MWsT setups, and values as high as 6% when the temperature fit is not taken into account. In relation to the systematic uncertainty profiles (Fig. 3e), they range from 3% in the UTLS and middle stratosphere (around 30 km) to 5% in the upper stratosphere. As with statistical errors, the temperature contribution decreases when including the atmospheric temperature profile in the retrieval, leading to smaller systematic errors in O<sub>3</sub> TCs, as mentioned above.

The error estimation presented here assumes the same set of uncertainty values for all setups, which is representative of the IFS 120/HR instrument in the period 2005–2008 (Table B1 in Appendix B). However, some error sources do strongly depend on instrument status (particularly the ILS

function, solar pointing, and measurement noise), affecting the total uncertainty budget. In order to account for the different quality periods of the IZO FTIR instruments, the uncertainty analysis for different sets of error values is included in Appendix B.

To summarise, 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<sub>3</sub> 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 is 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<sub>3</sub> retrievals. Finally, although the narrow micro-window setups provide very consistent results, the 5MWsT setup has been theoretically shown to be superior for the typical O<sub>3</sub> concentrations observed at tropical and subtropical latitudes.

## 4 Comparison to reference observations

### 4.1 FTIR and Brewer ozone total columns

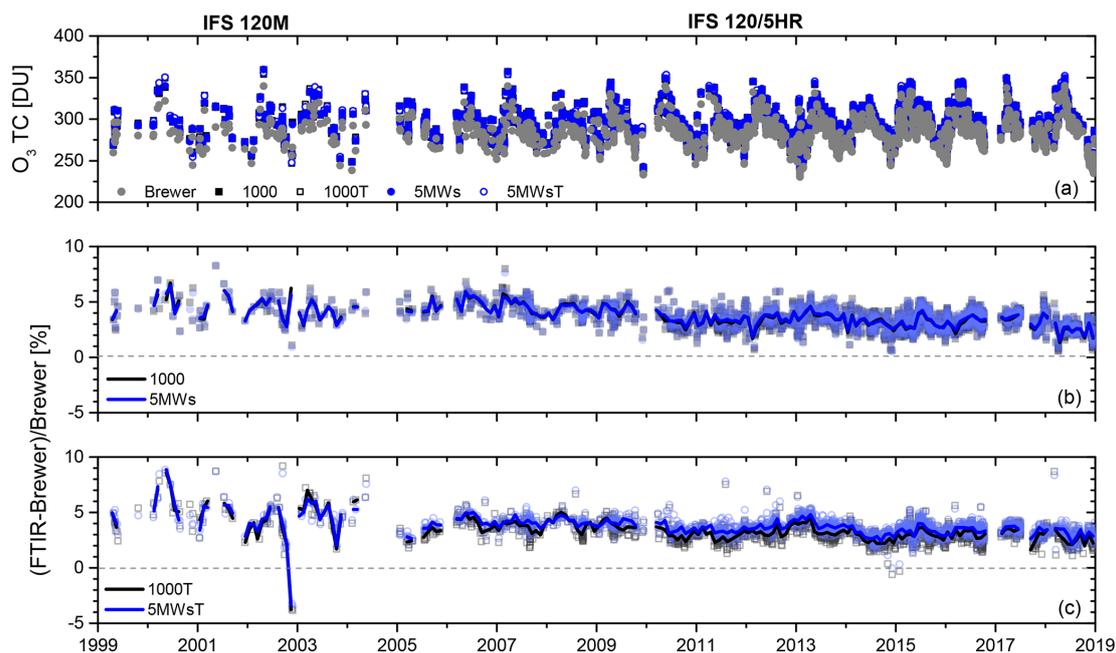
The performance of each of the six FTIR O<sub>3</sub> retrieval strategies has been assessed by comparison with coincident NDACC Brewer O<sub>3</sub> TC data. In order to mitigate the influence of the O<sub>3</sub> intra-day variations on comparisons, only FTIR and Brewer measurements within a temporal coincidence of 5 min have been paired, which makes a total of 2231 coincidences between 1999 and 2018.

Figure 4 displays the time series of the Brewer observations together with four examples of FTIR retrievals (for simplicity, only the setups 1000/1000T and 5MWs/5MWsT are depicted), as well as the time series of corresponding relative differences (RD, FTIR–Brewer). The temporal O<sub>3</sub> TCs variations are in general reproduced well by all FTIR products. However, this figure makes the difference in performance of the two FTIR instruments evident: while the RD values of the IFS 120/5HR instrument are very stable over time, the IFS 120M instrument exhibits more erratic behaviour. Besides the greater variability of the IFS 120M and the switching of instruments in 2005, the most remarkable feature is a discontinuity detected at the beginning of 2010 by a non-parametric change-point test (Lanzante, 1996) (at the 99 % confidence level). The systematic jump is  $\sim 1.1$  % for the setups without temperature fitting, and is partially corrected by retrieving the temperature:  $\sim 0.7$  %. The change point, already reported by García et al. (2014), is likely due to modifications of the IFS 120/5HR spectrometer (failure of the interferometer's scanner motor and its subsequent replacement). Another change point was detected by Lanzante's approach around the beginning of 2014, but was less intense. It is worth highlighting that both discontinuities were also detected in the differences

of O<sub>3</sub> TCs retrieved from the different FTIR setups, especially by those including the temperature fit. 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 4 also reveals that, although the scatter in the RD is significantly improved by the temperature retrieval, these strategies generally present more extreme values as compared to the Brewer data. The RD range between the 0.1st and 99.9th percentiles is 7.4 % and 7.0 % for the 1000 and 5MWs setups, 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 the 1000T and 5MWsT setups at the beginning of 2000 and at the end of 2002 (Fig. 4c) are not reproduced by the 1000 and 5MWs strategies (Fig. 4b). This pattern is consistently observed for all setups 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 that assume 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<sub>3</sub> retrieval procedure. Regarding Brewer, no temperature or pressure dependence is considered in the operational data processing (Redondas et al., 2014; Rimmer et al., 2018). In particular, the Brewer O<sub>3</sub> TCs are computed using the so-called effective O<sub>3</sub> cross sections throughout the atmosphere (Bass and Paur, 1985), corresponding to an O<sub>3</sub> effective height of 22 km and a fixed effective temperature of the O<sub>3</sub> layer of  $-45$  °C. These simplifications can produce systematic (seasonal dependence) and random errors (Redondas et al., 2014; Gröbner et al., 2021). In fact, at IZO, the effective temperature and O<sub>3</sub> height significantly differ from the values assumed by Brewer processing in winter months, when the extreme RD values are observed (Fig. 4c). Nevertheless, a more dedicated study would be desirable to deeply investigate the causes driving these anomalous values.

When analysing in detail the intercomparison results (Table 2), it is confirmed that, first, the more refined setups using narrow micro-windows offer the best performance (especially 5MWs/5MWsT) independently of the treatment of the atmospheric temperature profile, and second, the effect of the simultaneous temperature fit on the FTIR O<sub>3</sub> quality depends on the instrumental stability. The agreement between the FTIR and Brewer observations significantly worsens for the more unstable IFS 120M spectrometer when the temperature fit is included in the retrieval procedure (the largest median bias and scatter and the 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<sub>3</sub> products by



**Figure 4.** Summary of the FTIR–Brewer comparison from 1999 to 2018. **(a)** Time series of O<sub>3</sub> TCs (DU) as observed by Brewer and FTIR 1000/1000T and 5MWs/5MWsT setups. **(b)** Time series of relative differences (RD) for the setups 1000 and 5MWs, which were calculated as  $RD(\%) = 100 \times (O_3TC_X - O_3TC_Y) / O_3TC_Y$ , where  $X$  and  $Y$  refer to FTIR and Brewer, respectively. **(c)** As for **(b)**, but for the setups 1000T and 5MWsT. Solid lines in **(b)** and **(c)** correspond to monthly medians.

**Table 2.** Summary of statistics for the FTIR–Brewer comparison for the setups 1000/1000T, 4MWs/4MWsT, and 5MWs/5MWsT: the median ( $M$ , in %) and standard deviation ( $\sigma$ , in %) of the relative differences and Pearson correlation coefficient ( $R$ ) of the direct comparison are shown for the periods 1999–2004, 2005–May 2008, and June 2008–2018 and for the entire time series (1999–2018). Also shown are the median of the theoretical TPE ( $M_{TPE}$ ), SE ( $M_{SE}$ ), and TE ( $M_{TE}$ ) uncertainties. The number of coincident FTIR–Brewer measurements is 93, 185, and 1892 for the three periods, respectively, and 2170 for the whole dataset. The strategies showing the best performance in terms of smallest  $\sigma$  and  $M_{TE}$  are highlighted in bold for each period.

Setup	1999–2004						2005–2008						2008–2018						1999–2018					
	$M$ , (%)	$\sigma$ , (%)	$R$ , (%)	$M_{TPE}$ , (%)	$M_{SE}$ , (%)	$M_{TE}$ , (%)	$M$ , (%)	$\sigma$ , (%)	$R$ , (%)	$M_{TPE}$ , (%)	$M_{SE}$ , (%)	$M_{TE}$ , (%)	$M$ , (%)	$\sigma$ , (%)	$R$ , (%)	$M_{TPE}$ , (%)	$M_{SE}$ , (%)	$M_{TE}$ , (%)	$M$ , (%)	$\sigma$ , (%)	$R$ , (%)	$M_{TPE}$ , (%)	$M_{SE}$ , (%)	$M_{TE}$ , (%)
1000	4.29	1.38	0.957	1.85	0.23	1.87	4.47	0.86	0.970	1.85	0.23	1.87	3.35	0.83	0.982	1.83	0.23	1.85	3.46	0.95	0.975	1.84	0.23	1.85
4MWs	4.28	1.36	0.959	1.79	0.19	1.80	4.49	0.85	0.971	1.78	0.19	1.80	3.34	0.82	0.982	1.77	0.19	1.78	3.45	0.93	0.976	1.77	0.19	1.78
5MWs	4.35	<b>1.32</b>	0.962	1.68	0.16	1.68	4.53	0.82	0.973	1.66	0.17	1.67	3.41	0.81	0.983	1.65	0.17	1.66	3.50	0.91	0.977	1.65	0.17	1.66
1000T	4.83	1.97	0.926	0.52	0.24	0.57	3.79	0.82	0.972	0.51	0.15	0.53	3.04	0.73	0.986	0.51	0.16	0.53	3.12	0.90	0.977	0.51	0.16	0.53
4MWsT	4.84	1.90	0.934	0.44	0.21	0.48	3.97	0.66	0.981	0.42	0.10	0.44	3.32	0.68	0.988	0.42	0.12	0.44	3.40	0.83	0.981	0.42	0.12	0.44
5MWsT	4.81	1.82	0.940	0.40	0.22	<b>0.44</b>	4.15	<b>0.63</b>	0.983	0.39	0.09	<b>0.40</b>	3.44	<b>0.67</b>	0.988	0.39	0.11	<b>0.40</b>	3.53	<b>0.81</b>	0.982	0.39	0.11	<b>0.40</b>

considerably reducing the dispersion and bias of the RD distributions. Thus, the best performance is obtained by those setups using narrow micro-windows, with a RD scatter of only  $\sim 0.6\%$ – $0.7\%$  (this increases up to  $\sim 0.8\%$  for the broader region setup) for the IFS 120/5HR instrument; while it is as high as 2% 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 (see Sects. 2.2 and 3.2.2 and Appendix B). In fact, as shown in Appendix B, total statistical errors of  $\sim 0.5\%$ – $1.5\%$  can be expected depending on the instrument status

(i.e. the instrumental degradation and solar pointing errors). Nonetheless, Table 2 also documents that the scatter found in the RD is noticeably lower than that predicted when the temperature fit is not considered, especially for the IFS 120/5HR instrument (see  $M_{TE}$  values). This fact could indicate that sources of uncertainty partially cancel each other out, which is not fully reproduced by the theoretical error assessment, and/or the possible overestimation of the assumed temperature uncertainty. As also shown in Appendix B, reducing the latter would contribute to reconciling the experimental and theoretical results. Note that the scatter values found for the IFS 120/5HR spectrometer can be used to derive a conserva-

tive value for the precision of the FTIR O<sub>3</sub> TC estimations, since they can be interpreted as the root of the sum of squares of the precision of the Brewer or FTIR instrument.

Regarding the systematic differences, a median bias of  $\sim 3\%$ – $5\%$  is obtained. Such discrepancies are consistent with previous studies (e.g. Schneider et al., 2008a; García et al., 2012; García et al., 2016), and are mainly attributed to inconsistencies between infrared and ultraviolet spectroscopic parameters (e.g. Piquet-Varrault et al., 2005; Gratien et al., 2010; Drouin et al., 2017; Tyuterev et al., 2019). In fact, as recently presented by Gordon et al. (2022), the most recent release of the HITRAN spectroscopic database (HITRAN 2020) improves the O<sub>3</sub> line intensities in the 1000 cm<sup>-1</sup> spectral band by applying a scaling factor of 3%. This correction agrees well with our theoretical uncertainty estimations and the overestimation found for the FTIR products.

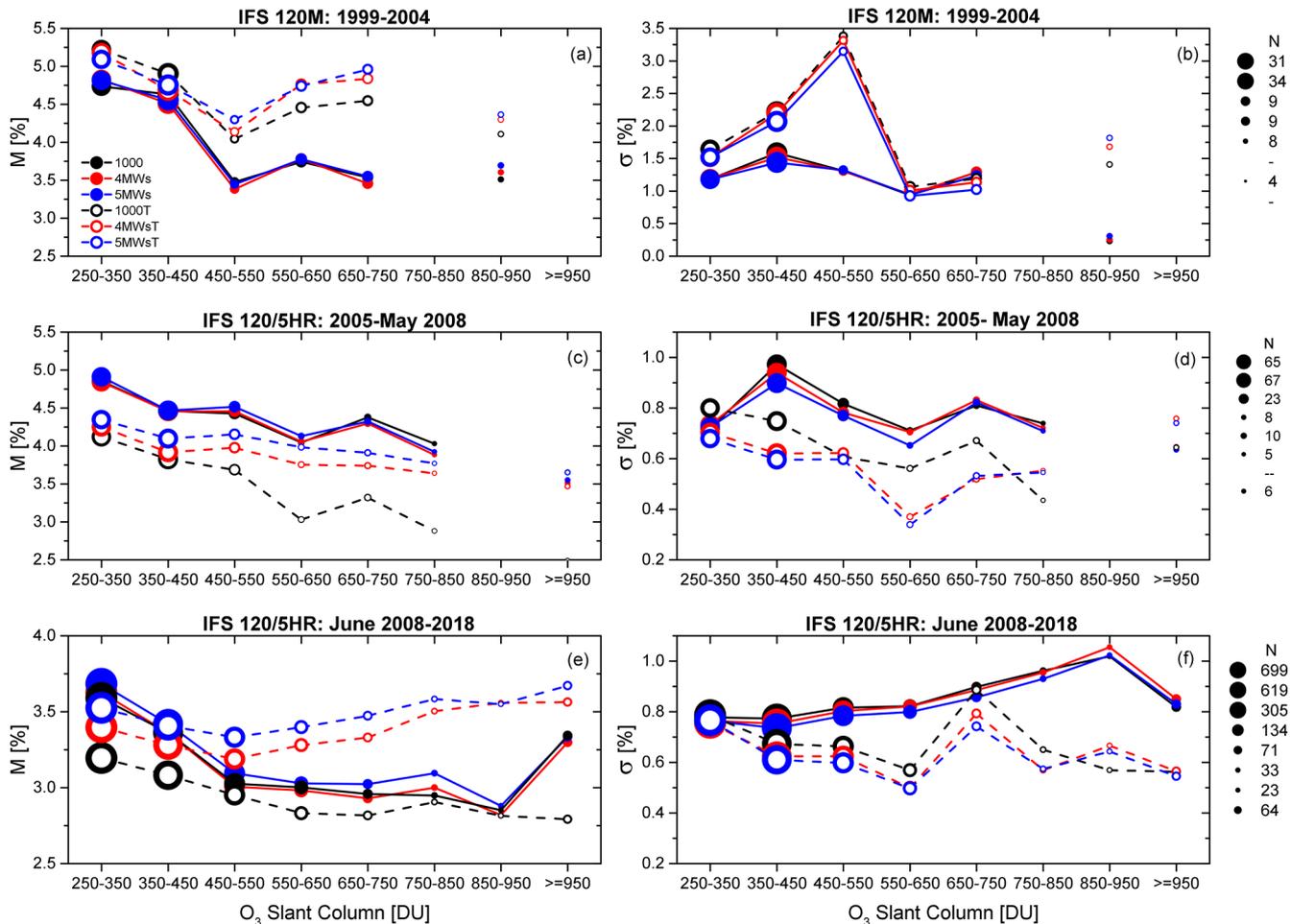
Similar conclusions can be reached in general when the comparison is performed as a function of O<sub>3</sub> signatures in the slant path (Fig. 5). The temperature fit improves the performance of the stable instrument and makes it worse for the more unstable instrument, independently of the O<sub>3</sub> SC range covered at IZO. In addition, the bias between FTIR and Brewer data decreases by  $\sim 1\%$  overall as the O<sub>3</sub> SCs increase for all setups (Fig. 5a, c, and e). This dependence can in part be accounted for the Brewer systematic uncertainties in the absolute calibration process, which are amplified for low O<sub>3</sub> SCs (up to 0.5%; Schneider et al., 2008a). However, for the more stable IFS 120/5HR period (Fig. 5e), the bias alters this behaviour for O<sub>3</sub> SCs beyond  $\sim 550$  DU for setups using narrow micro-windows when the simultaneous temperature retrieval is considered. This issue could be attributed to inconsistencies in the spectroscopic parameters at higher wavenumbers, which gain importance as O<sub>3</sub> concentrations increase, and is in line with the theoretical systematic inconsistency found between the 1000T and 4MWsT/5MWsT setups (Sect. 3.2.2). However, the number of FTIR–Brewer coincidences at IZO is rather small for O<sub>3</sub> SCs greater than 550 DU (i.e. less than  $\sim 20\%$  of the Brewer data in the 1999–2018 period); therefore, a more robust dataset would be recommended to better understand what drives this different pattern.

As pointed out by the theoretical uncertainty analysis, statistical errors are expected to increase with O<sub>3</sub> SCs for all setups. This can be seen in the scatter of the RD for the IFS 120/5HR periods (Fig. 5d and f) when the temperature fit is not applied (i.e. the scatter of the RD increases by 0.2% in the 2008–2018 period at larger O<sub>3</sub> SCs). However, the inter-comparison results seem not to exhibit a similar dependence when the temperature retrieval is considered, as expected. At larger O<sub>3</sub> SCs, including this fit can lead to differences of 0.2% compared to when this fit is not included. This could be attributed to an underestimation of the ILS and/or baseline errors in the theoretical assessment: the ILS contribution to the total uncertainty budget decreases as the O<sub>3</sub> SCs increase and becomes more important when the temperature fit is applied

(see Fig. B1). The same behaviour is found for the baseline error (data not shown). Hence, an increment in the assumed uncertainties for these two error sources when the temperature is included in the retrieval procedure could partially reduce discrepancies between the theoretical and experimental assessments. Note that while the 1000/1000T setups provide the most accurate O<sub>3</sub> TCs with respect to Brewer data for the whole O<sub>3</sub> SC range for both instruments, the 5MWs/5MWsT setups offer the most precise O<sub>3</sub> TCs. This result further corroborates that the broad region seems to be less sensitive to the improvement generated by the temperature retrieval.

The long-term FTIR time series used in this study allows us to investigate not only 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 of greater than 0.99 for all retrieval strategies considered (Fig. 6a and b). However, the bias between both techniques depends on the O<sub>3</sub> amounts, leading to a seasonal effect on RD, which is likely due to the fact that the Brewer and FTIR products exhibit different responses to O<sub>3</sub> seasonal variations. On the one hand, the FTIR sensitivity is strongly anti-correlated with the O<sub>3</sub> SC annual cycle: the lower the O<sub>3</sub> amounts, the less saturated the O<sub>3</sub> absorption lines. This results in minimum (maximum) DOFS in winter (spring/summer) (García et al., 2012). On the other hand, the different treatments of the atmospheric temperature and the O<sub>3</sub> vertical distribution in the Brewer data processing and the FTIR data processing also generate seasonal artefacts, as stated above for the extreme RD values. In fact, including the temperature retrieval significantly modifies the RD seasonal patterns, as observed in Fig. 6c and d.

For those approaches without a simultaneous temperature fit (Fig. 6c), the RD annual cycle seems to follow the typical O<sub>3</sub> 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 the subtropical tropopause and the annual cycle of O<sub>3</sub> photochemical production associated with tropical insolation (e.g. García et al., 2014; García et al., 2021). Hence, a significant correlation between the averaged RD and O<sub>3</sub> TC annual cycles for all setups is found, with Pearson correlation coefficients ranging from 0.68 to 0.84 for the 1000 and 5MWs strategies, respectively. This relationship drops to correlation values between 0.32 and 0.60 for the 1000T and 5MWsT setups, respectively. However, in return, a seasonal dependence on the upper-stratospheric temperature is detected (Fig. 6f displays, as an example, the annual cycle of averaged temperature at 39 km from the NCEP database along with those retrieved from the FTIR setups). The correlation between the averaged annual cycles of the upper-stratospheric temperature and RD is  $\sim 0.70$  and  $\sim 0.90$  for the broad and narrow micro-window setups, respectively, when the temperature fit is included; it is limited to between 0.46 and 0.61 when the temperature re-



**Figure 5.** Median and standard deviation of relative differences (RD, FTIR–Brewer) with respect to the Brewer  $O_3$  slant column (DU) for the periods 1999–2004, 2005–May 2008, and June 2008–2018. Panels (a), (c), and (e) show median RD ( $M$ , in %) values for the three periods, respectively, and (b), (d), and (f) show the same but for the standard deviation of the 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 better visualisation, a scale factor of 3, 2, and 1 was applied to the dotted area for the periods 1999–2004, 2005–May 2008, and June 2008–2018, respectively.

retrieval is not considered. Note that a subtle relationship with the temperature in the middle or lower stratosphere (e.g. at 29 km in Fig. 6e) is found. Additionally, it has been found that the RD seasonal amplitudes are augmented overall by the temperature retrieval. 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 is modified by less than 0.05 % for the 4MWs/5MWs setups.

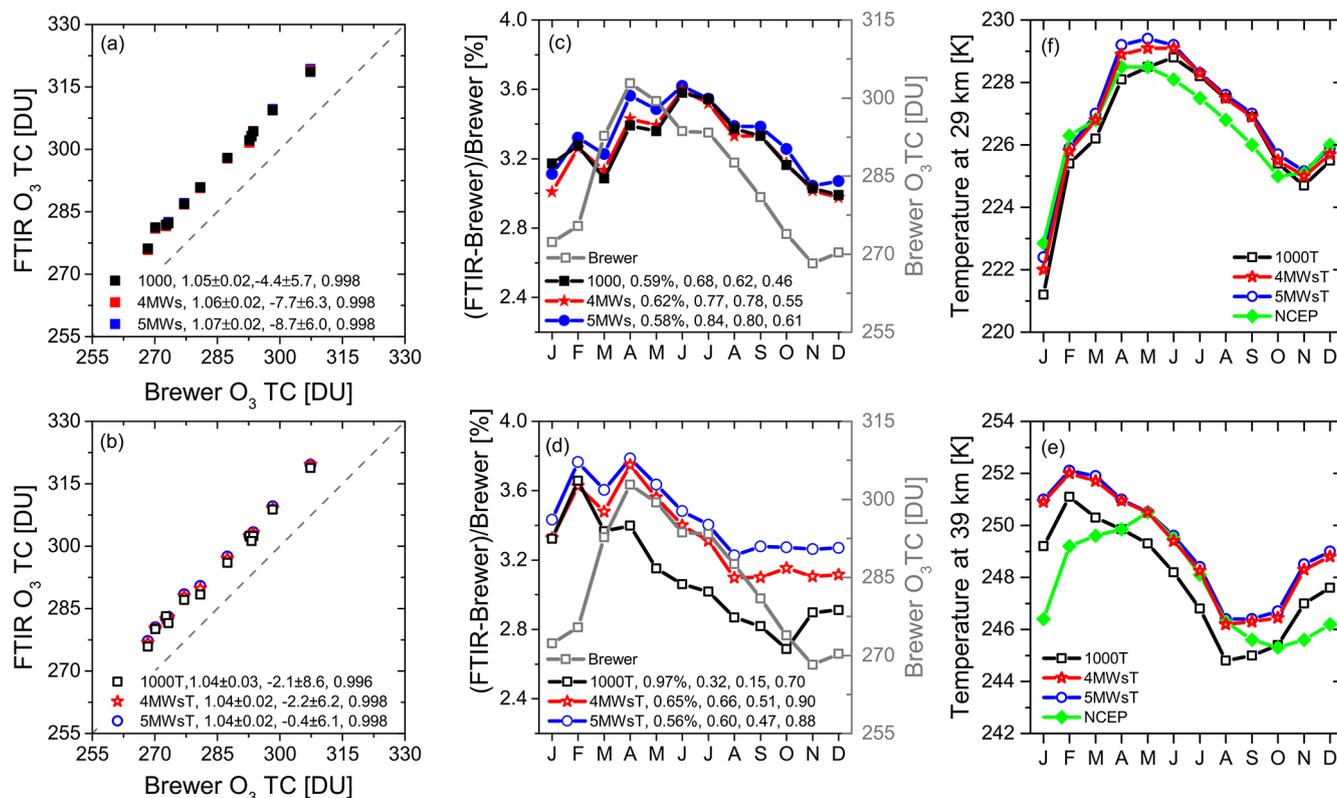
#### 4.2 FTIR and ECC ozone vertical profiles

In order to evaluate the influences of the six retrieval strategies on the  $O_3$  vertical distribution, Fig. 7 displays the vertical profiles of the relative differences between FTIR and ECC sondes for the three periods considered, while Table 3 summarises the comparison for the  $O_3$  layers that are sufficiently detectable by the FTIR system, i.e. the partial col-

umn (PC) at 2.37–13, 12–23, and 22–29 km (the DOFS for all these layers is typically larger than 1). 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 ( $x_{\text{ECC}}$ ) were vertically degraded ( $\hat{x}_{\text{ECC}}$ ) by applying the averaging kernels obtained in the FTIR  $O_3$  retrieval procedure (Rodgers, 2000) as follows:

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

where  $x_a$  is the a priori  $O_3$  VMR profiles. The ECC smoothing allows, on the one hand, the limited sensitivity of FTIR data to be taken into account and, on the other hand, the ef-

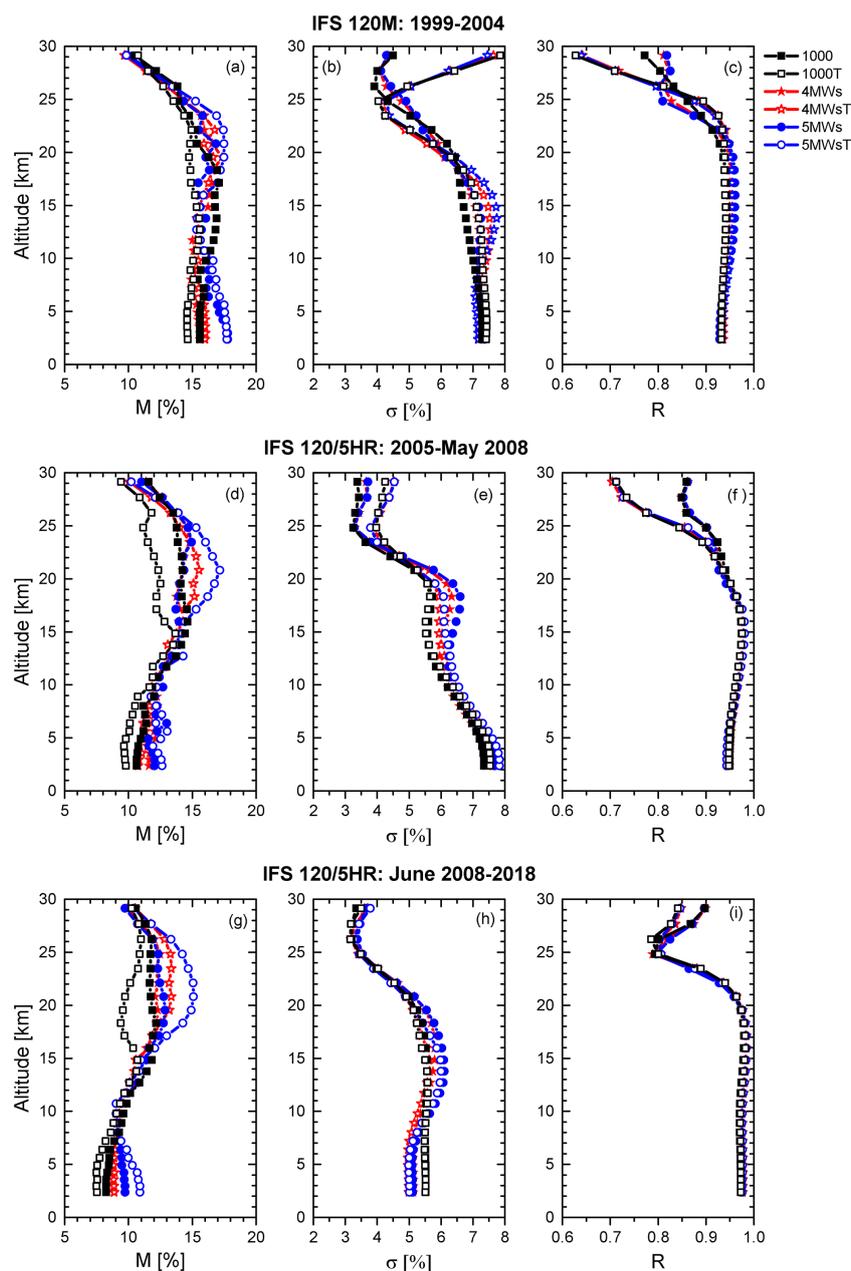


**Figure 6.** Summary of the FTIR–Brewer comparison at a seasonal scale for the 2009–2018 period. (a, b) Scatter plots of the Brewer and FTIR O<sub>3</sub> TCs for the setups 1000/4MWs/5MWs and 1000T/4MWsT/5MWsT, respectively. (c, d) Averaged annual cycles of the Brewer O<sub>3</sub> TC and RD for the setups 1000/4MWs/5MWs, and 1000T/4MWsT/5MWsT, respectively. (e, f) Averaged annual cycles of the atmospheric temperature at 29 km and 39 km, respectively, retrieved from the setups 1000T, 4MWsT, and 5MWsT, as well as the NCEP database. The legends in (a) and (b) display the slope, offset, and Pearson correlation coefficient of the least-squares fit, and the legends in (c) and (d) show the amplitude of the RD annual cycle (in %), the Pearson correlation coefficient between the RD and Brewer O<sub>3</sub> TC annual cycles, and the annual cycle of atmospheric temperature (at 29 and 39 km, respectively).

facts of the different strategies on retrieved O<sub>3</sub> profiles to be directly 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<sub>3</sub> retrievals for computing  $\hat{x}_{\text{ECC}}$ . Finally, the temporal collocation window between FTIR and ECC sondes extends to  $\pm 3$  h around the sonde launch (typically at 12:00 UT) 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  $\sim 30$  km are clearly discernible (troposphere, UTLS, and middle stratosphere; recall Fig. 3). In particular, beyond the UTLS region, the influence of ILS uncertainties on the retrieved O<sub>3</sub> profiles becomes important with altitude, since the full width at half maximum of the narrow O<sub>3</sub> absorption lines and the ILS function becomes comparable. For all setups and periods, the simultaneous temperature fit is found to worsen the agreement between FTIR and

ECC sondes at higher altitudes (see the standard deviation profiles in Fig. 7b, e, and h). However, as the instrument becomes better aligned and more stable over time, the effect of this cross-interference becomes less significant, until no noticeable differences are observed for the 2008–2018 period. For example, the scatter at 29 km for the FTIR–ECC comparison only changes from 3.6 % to 3.8 % for the 5MWs and 5MWsT setups, 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, overall, the best agreement is found for those setups without temperature profile fitting. This may indicate that the negative effect of the ILS uncertainties and measurement noise prevails over the improvement attributed to the temperature fit. The same pattern is documented for the tropospheric O<sub>3</sub> concentrations, even though the differences among retrieval strategies are not as significant as those at higher altitudes. Note that these scatter values agree well with the expected uncertainty for ECC sondes ( $\sim 5\%$ – $15\%$ ) and with the FTIR theoretical error estimation (recall Sect. 3.2.2), as well as with previous works



**Figure 7.** Summary of the FTIR-smoothed ECC comparison for the periods 1999–2004, 2005–May 2008, and June 2008–2018. Panels (a), (d), and (g) display the vertical profiles of median ( $M$ ) RD (FTIR–ECC, in %) for the three periods, respectively. (b, e, h) As for (a), (d), and (g), but for the standard deviation of the RD distribution ( $\sigma$ , in %). (c, f, i) As for (a), (d), and (g), but for the 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.

(Schneider et al., 2008b; García et al., 2012; Duflot et al., 2017, and references therein). As stated in those 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 (Fig. 7a, d, and g). While all setups consistently show larger biases up to the UTLS region for the IFS 120M, the differences between the instruments are minimised beyond the middle stratosphere. The median bias varies from  $\sim 17\%$  to  $\sim 10\%$  at 5 km for the 5MWs/5MWsT in the 1999–2004 and 2008–

**Table 3.** Same as Table 2, but for the FTIR–smoothed ECC comparison for the O<sub>3</sub> partial columns computed for 2.37–13, 12–23, and 22–29 km. The number of coincident FTIR–ECC measurements is 56, 49, and 167 for the three periods, respectively, and 272 for the whole dataset. The strategies showing the best performance, in terms of smallest  $\sigma$  and  $M_{TE}$ , are highlighted in bold for each period.

Setup	1999–2004			2005–2008			2008–2018			1999–2018		
	<i>M</i> , $\sigma$ , <i>R</i> , $M_{TPE}$ , $M_{SE}$ , $M_{TE}$ (%), (%), –, (%), (%), (%)	<i>M</i> , $\sigma$ , <i>R</i> , $M_{TPE}$ , $M_{SE}$ , $M_{TE}$ (%), (%), –, (%), (%), (%)	<i>M</i> , $\sigma$ , <i>R</i> , $M_{TPE}$ , $M_{SE}$ , $M_{TE}$ (%), (%), –, (%), (%), (%)	<i>M</i> , $\sigma$ , <i>R</i> , $M_{TPE}$ , $M_{SE}$ , $M_{TE}$ (%), (%), –, (%), (%), (%)	<i>M</i> , $\sigma$ , <i>R</i> , $M_{TPE}$ , $M_{SE}$ , $M_{TE}$ (%), (%), –, (%), (%), (%)	<i>M</i> , $\sigma$ , <i>R</i> , $M_{TPE}$ , $M_{SE}$ , $M_{TE}$ (%), (%), –, (%), (%), (%)						
FTIR–ECC at 2.37–13 km												
1000	15.86, 7.08, 0.934, 1.14, 6.34, 6.44	11.44, 6.58, 0.956, 1.09, 5.93, 6.03	9.08, 5.46, 0.970, 1.06, 5.68, 5.78	10.63, 6.61, 0.953, 1.07, 5.76, 5.86								
4MWs	15.08, 7.11, 0.937, 1.26, 5.76, 5.91	12.08, <b>6.54</b> , 0.958, 1.28, 6.32, 5.48	9.38, 5.10, 0.974, 1.25, 5.17, 5.33	10.62, 6.46, 0.956, 1.25, 5.24, 5.39								
5MWs	16.24, 7.09, 0.938, 1.01, 5.06, 5.16	12.30, 6.64, 0.957, 1.06, 4.75, 4.87	9.72, 5.13, 0.974, 1.02, 4.61, 4.72	10.89, 6.56, 0.956, 1.03, 4.66, 4.78								
1000T	15.05, 7.30, 0.934, 0.79, 6.05, 6.10	10.39, 6.73, 0.954, 0.63, 5.64, 5.67	8.12, 5.45, 0.971, 0.62, 5.46, 5.50	10.00, 6.67, 0.953, 0.64, 5.53, 5.57								
4MWsT	15.56, 7.22, 0.938, 0.91, 5.67, 5.74	11.56, 6.76, 0.957, 0.79, 5.23, 5.29	9.26, <b>4.94</b> , 0.975, 0.77, 5.10, 5.16	10.54, <b>6.36</b> , 0.958, 0.79, 5.16, 5.22								
5MWsT	16.93, <b>7.05</b> , 0.940, 0.69, 4.99, <b>5.03</b>	12.76, 6.82, 0.956, 0.60, 4.68, <b>4.72</b>	10.11, 4.99, 0.976, 0.59, 4.53, <b>4.57</b>	11.20, 6.42, 0.958, 0.60, 4.59, <b>4.63</b>								
FTIR–ECC at 12–23 km												
1000	17.30, 5.76, 0.914, 0.81, 2.08, 2.24	16.01, <b>4.54</b> , 0.946, 0.75, 1.75, 1.91	13.23, <b>4.70</b> , 0.959, 0.76, 1.71, 1.88	14.59, 5.23, 0.944, 0.76, 1.74, 1.91								
4MWs	17.91, 5.58, 0.928, 0.50, 1.77, 1.84	15.99, 4.86, 0.943, 0.47, 1.53, 1.60	13.41, 4.82, 0.961, 0.47, 1.53, 1.61	14.79, 5.29, 0.947, 0.47, 1.55, 1.63								
5MWs	17.89, <b>5.43</b> , 0.939, 0.52, 1.65, <b>1.73</b>	16.03, 4.94, 0.943, 0.59, 1.47, <b>1.59</b>	13.72, 4.81, 0.962, 0.59, 1.47, <b>1.59</b>	14.91, 5.21, 0.951, 0.58, 1.48, <b>1.60</b>								
1000T	16.83, 5.64, 0.921, 0.66, 1.95, 2.06	15.19, 4.59, 0.945, 0.62, 1.63, 1.75	12.13, 4.75, 0.961, 0.62, 1.63, 1.74	13.72, 5.27, 0.946, 0.62, 1.65, 1.77								
4MWsT	17.25, 5.61, 0.930, 0.48, 1.82, 1.88	16.44, 4.65, 0.947, 0.45, 1.58, 1.65	14.18, 4.73, 0.963, 0.44, 1.60, 1.66	15.30, 5.12, 0.951, 0.45, 1.62, 1.68								
5MWsT	17.48, 5.72, 0.935, 0.38, 1.76, 1.81	16.91, 4.67, 0.948, 0.38, 1.56, 1.60	15.07, 4.77, 0.963, 0.37, 1.56, 1.61	16.02, <b>5.11</b> , 0.953, 0.37, 1.58, 1.62								
FTIR–ECC at 22–29 km												
1000	15.92, 3.71, 0.820, 2.00, 2.87, 3.51	16.23, <b>2.87</b> , 0.888, 2.17, 2.58, 3.38	13.94, 3.17, 0.756, 2.08, 2.56, 3.30	14.76, 3.50, 0.779, 2.08, 2.59, 3.33								
4MWs	16.09, 4.01, 0.793, 2.32, 2.71, 3.59	16.67, 2.88, 0.888, 2.66, 2.47, 3.63	14.16, 3.21, 0.767, 2.54, 2.53, 3.59	14.80, 3.64, 0.773, 2.53, 2.53, 3.59								
5MWs	16.00, 4.10, 0.783, 2.55, 2.78, 3.79	16.43, 2.87, 0.892, 2.91, 2.55, 3.88	14.06, 3.23, 0.777, 2.80, 2.62, 3.84	14.79, 3.68, 0.776, 2.79, 2.63, 3.84								
1000T	16.80, <b>3.54</b> , 0.866, 0.88, 3.10, 3.22	15.56, 3.12, 0.856, 0.78, 2.80, 2.91	13.88, 3.22, 0.735, 0.76, 2.77, 2.89	14.51, 3.55, 0.778, 0.77, 2.80, 2.90								
4MWsT	16.28, 3.56, 0.864, 0.77, 3.01, 3.11	15.61, 3.36, 0.844, 0.67, 2.63, 2.71	14.45, <b>3.05</b> , 0.774, 0.67, 2.69, 2.77	14.96, <b>3.43</b> , 0.801, 0.68, 2.71, 2.79								
5MWsT	16.31, 3.67, 0.856, 0.74, 2.94, <b>3.03</b>	15.28, 3.53, 0.829, 0.66, 2.59, <b>2.67</b>	14.47, 3.12, 0.775, 0.65, 2.65, <b>2.73</b>	15.06, 3.50, 0.798, 0.66, 2.67, <b>2.75</b>								

2018 periods, respectively, while the bias is  $\sim 10\%$  at 29 km for the same configurations for both periods.

In summary, considering the integrated PCs (which are less dependent on the FTIR vertical sensitivity) and the 2008–2018 period as a reference (due to better instrumental alignment and more FTIR–ECC coincidences), the best overall performance is documented for the setups using narrow micro-windows and simultaneous temperature fits in the troposphere and stratosphere regions (Table 3). At UTLS altitudes, where O<sub>3</sub> is particularly variable, the broad micro-window strategy seems to provide the best agreement with respect to ECC data. Nevertheless, the differences among strategies lie within the respective error confidence intervals, and so 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  $\sim 30$  km. However, compensations among the ILS, temperature, measurement errors, and O<sub>3</sub> vertical distribution should occur at higher altitudes, meaning that the usage of narrow micro-windows (and a temperature fit) clearly provides 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, so other measurement techniques, such as microwave or lidar O<sub>3</sub> profiles, would be of great use for further completing the quality assessment.

## 5 Summary and conclusions

Accurate ozone (O<sub>3</sub>) 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<sub>3</sub> products from ground-based FTIR spectrometry, with the aim of providing an improved O<sub>3</sub> 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 have been utilised. The 20-year time series of O<sub>3</sub> observations allowed us to, on the one hand, assess the quality and long-term consistency of the different FTIR O<sub>3</sub> products and, on the other hand, evaluate the influence of instrumental status on the O<sub>3</sub> retrievals.

The quality of the FTIR O<sub>3</sub> products improves as the retrieval strategies become more refined by considering O<sub>3</sub> absorption lines in specific narrow micro-windows (between 991 and 1014 cm<sup>−1</sup>) instead of using the traditional broad spectral region (between 1000 and 1005 cm<sup>−1</sup>). Approaches using narrow micro-windows proved, both theoretically and experimentally, to be superior due to their greater vertical sensitivity, smaller expected uncertainties, and better agreement with respect to independent data. Optimal selection of the spectral O<sub>3</sub> micro-windows enhanced the precision of FTIR O<sub>3</sub> TCs by  $\sim 0.1\%$ – $0.2\%$  with respect to the coincident NDACC Brewer observations taken as reference,

leading to a conservative precision of  $\sim 0.6\%$ – $0.7\%$  for the FTIR products. But, at the same time, they were shown to be consistent with the standard NDACC setup (i.e. no important biases were found between the different retrieval strategies).

In addition, independently of the O<sub>3</sub> absorption lines used, simultaneous atmospheric temperature retrieval was found to be a very useful tool for O<sub>3</sub> monitoring by ground-based NDACC FTIR systems. The scatter with respect to the Brewer data was found to be reduced up to  $\sim 0.2\%$  when applying a temperature fit for those strategies that also use narrow O<sub>3</sub> absorption lines. However, this improvement can only be reached provided the FTIR instrument is properly characterised and stable over time (e.g. the IFS 120/5HR spectrometer). For more unstable instruments, such as the IFS 120M, the inclusion of the atmospheric temperature fit in the O<sub>3</sub> retrieval procedure may not be recommendable, since it worsens the quality of FTIR O<sub>3</sub> products due to the increase in cross-interference with instrumental performance. The broad  $1000\text{ cm}^{-1}$  region seems to be most sensitive to this effect. Another fact that strongly distinguishes the broad and narrow setups is the presence of strong H<sub>2</sub>O absorption lines in the  $1000\text{ cm}^{-1}$  region, which could be critical for humid FTIR sites if the H<sub>2</sub>O cross-interference is not properly taken into account. In this sense, using one-step or two-step retrieval strategies (retrieving H<sub>2</sub>O and O<sub>3</sub> in the same step or in two separate steps, respectively) has been found to be valid and to provide consistent results.

Regarding the vertical O<sub>3</sub> distribution, the significant cross-interference between the O<sub>3</sub> and temperature profiles and the instrumental status causes the best-performing setup to depend on the altitude range considered. The best overall performance is documented for the setups using narrow micro-windows and simultaneous temperature fits in the troposphere and stratosphere regions, while at tropopause altitudes, the broad micro-window strategy seems to provide the best agreement with respect to ozonesonde data.

The effects of the most influential settings on the FTIR O<sub>3</sub> retrieval procedure have been examined in this paper. Nevertheless, there is great potential for further improving the precision and accuracy of FTIR O<sub>3</sub> products, as well as their harmonisation within the NDACC IRWG community (comprising instruments and retrieval strategies). Additional efforts could be made with the treatment of the instrumental response through the evaluation of the instrumental line shape function in a consistent manner, given its important effects on O<sub>3</sub> retrievals. In addition, testing the proposed O<sub>3</sub> setups at different NDACC FTIR stations (under different humidity conditions, latitudes, altitudes, etc.) would indeed motivate the NDACC FTIR community to revise the standard O<sub>3</sub> retrieval strategy. Improved O<sub>3</sub> monitoring could help to estimate the small expected signal of the recovery or decline in O<sub>3</sub> concentrations more precisely, for both integrated total columns and vertical distributions, at a global scale. This is particularly challenging in those regions where O<sub>3</sub> concentrations are less variable, such as at tropical and subtropi-

cal latitudes. Furthermore, there will be new opportunities to better understand the different and the main impacts of O<sub>3</sub> on the earth's climate system, improving their representation in current global climate models and, thus, knowledge of their long-term evolution.

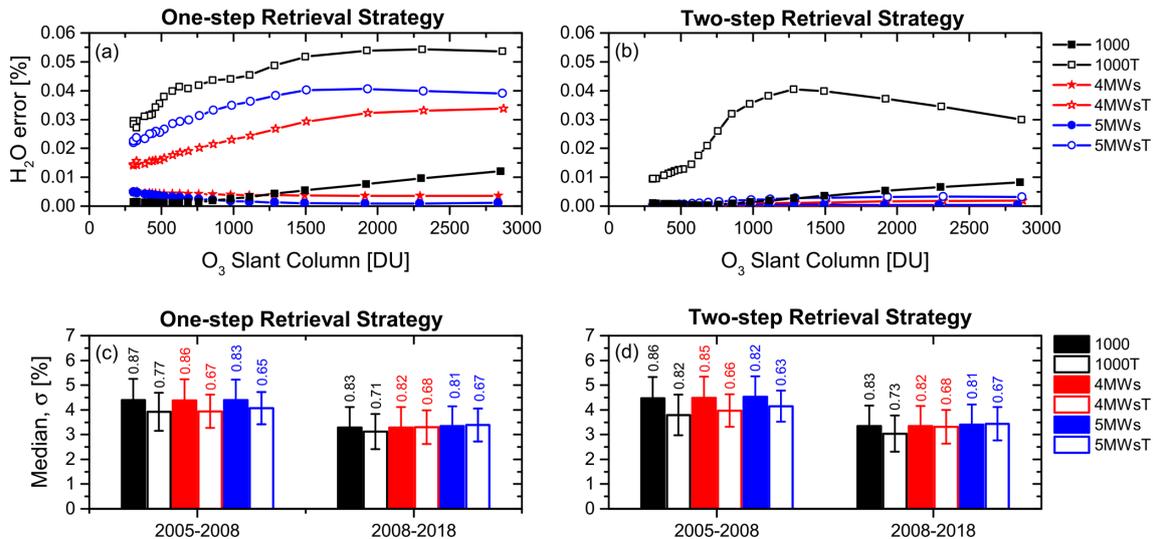
## Appendix A: Water vapour treatment

This appendix addresses the impact of the treatment of H<sub>2</sub>O on O<sub>3</sub> retrievals for all O<sub>3</sub> setups comparing the one-step and two-step retrieval strategies, where:

- The one-step approach refers to the simultaneous retrieval of the H<sub>2</sub>O and O<sub>3</sub> profiles using a Tikhonov–Phillips slope constraint for both gases, so that the micro-window of  $896.4\text{--}896.6\text{ cm}^{-1}$  is added for better H<sub>2</sub>O determination (as done at the NDACC FTIR Lauder and Wollongong sites in Vigouroux et al., 2015).
- The two-step approach refers to the strategy followed in the current paper and explained in detail in Sect. 3.1, where H<sub>2</sub>O a priori profiles, previously retrieved from dedicated H<sub>2</sub>O micro-windows for each spectrum, are then scaled in the O<sub>3</sub> retrieval.

Figure A1 illustrates the theoretical assessment of the H<sub>2</sub>O cross-interference for both retrieval strategies (García et al., 2014), where it can be seen that the H<sub>2</sub>O interfering error is noticeable, but not critical, given the especially dry conditions at IZO (the water vapour mixing ratio is less than 0.06 %) (García et al., 2021). Nevertheless, it has been found that the H<sub>2</sub>O interference strongly depends on the micro-windows used for the O<sub>3</sub> retrievals (the highest impact is observed for the 1000 spectral region, as expected from Fig. 2), as well as on the treatment of the atmospheric temperature profile. For all setups, the one-step and two-step approaches provide consistent results when a simultaneous temperature fit is not included, so it is valid to use either to correctly minimise the H<sub>2</sub>O interference. Nonetheless, provided the documented improvement of the temperature retrieval is pursued, the two-step strategy ought to be used. In this sense, the two-step strategy drastically reduces the H<sub>2</sub>O interfering error for those setups using narrow micro-windows when the simultaneous temperature fit is included (4MWsT/5MWsT), leading to expected errors in the O<sub>3</sub> total columns smaller than 0.01 %. The interfering effect of H<sub>2</sub>O also drops for the 1000 spectral region, but to a lesser extent, given the presence of important H<sub>2</sub>O absorption lines in that region (recall Fig. 2). This should be especially taken into account for FTIR stations located in humid environments.

A comparison to Brewer observations (Fig. A1c and d) also corroborates the theoretical results. It is worth highlighting the fact that the differences found between the two strategies are in excellent agreement with the estimated H<sub>2</sub>O interfering error values (Fig. A1a and b).



**Figure A1.** Summary of the impact of the treatment of H<sub>2</sub>O on O<sub>3</sub> retrievals for all O<sub>3</sub> setups. (a, b) H<sub>2</sub>O cross-interference error in the O<sub>3</sub> total column (%) using the one-step and two-step retrieval strategies, respectively, as a function of O<sub>3</sub> slant column (DU) for the measurements of Fig. 3. (c, d) Statistics for the FTIR–Brewer relative differences using the one-step and two-step retrieval strategies, respectively, for the periods 2005–May 2008 and June 2008–2018. Median ( $M$ , in %) and standard deviation ( $\sigma$ , presented as error bars and numerals in %) values are shown.

## Appendix B: Uncertainty analysis

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

$$(\hat{x} - x) = (A - I)(x - x_a) + GK_p(p - \hat{p}) + G\epsilon, \quad (\text{B1})$$

where  $G$  represents the gain matrix,  $K_p$  is the sensitivity matrix to the model parameters,  $I$  is the identity matrix, and  $A$  is the averaging kernel matrix.

The first term of Eq. (B1) refers to the smoothing error, which has been calculated as  $(A - I)S_{aO_3}(A - I)^T$ . The  $S_{aO_3}$  matrix is the O<sub>3</sub> 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 ( $S_{x,\epsilon}$ ) is analytically calculated by

$$S_{x,\epsilon} = GS_{y,\epsilon}G^T, \quad (\text{B2})$$

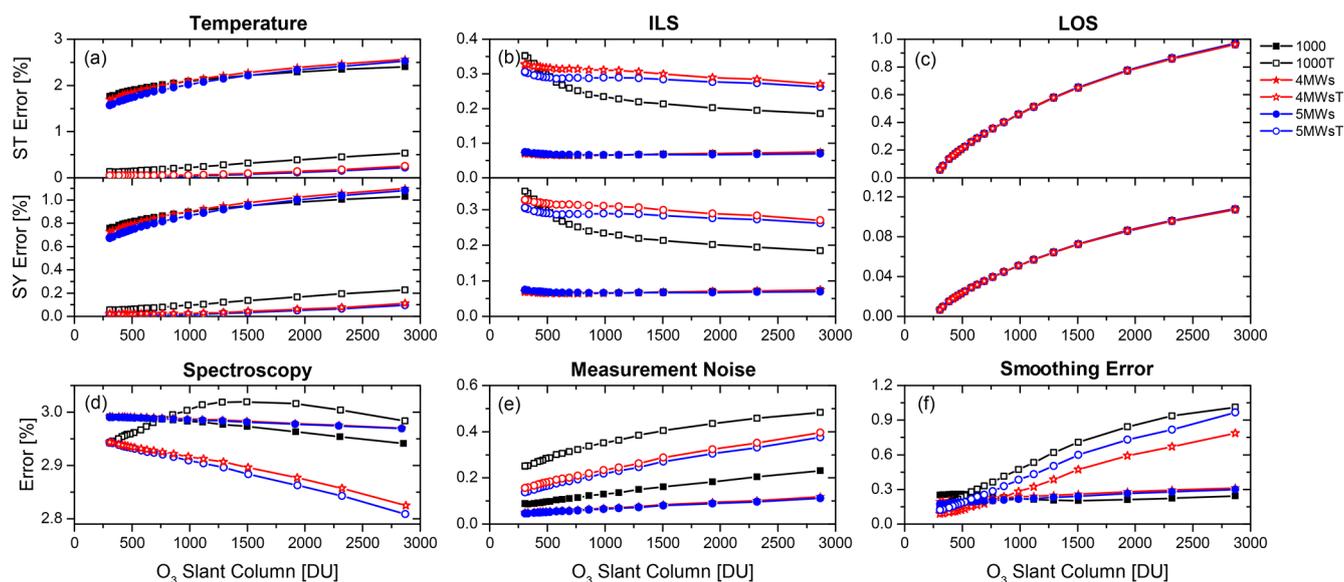
where  $S_{y,\epsilon}$  is the covariance matrix for measurement noise in the measurement.

The error contribution of the model parameters  $p$  can be analytically estimated through the respective error covariance matrices  $S_{x,p}$ :

$$S_{x,p} = GK_pS_pK_p^TG^T, \quad (\text{B3})$$

where  $S_p$  is the covariance matrix of the uncertainties  $\Delta p$ . In the current paper,  $S_p$  is estimated considering error sources, values, and the partitioning between random and systematic contributions listed in Table B1. 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; Gordon et al., 2022, and references therein). The statistical and systematic contributions of total parameter errors (TPE, displayed in Fig. 3) are then calculated as the square root of the sum of the squares of all statistical and systematic errors considered, respectively. Note that measurement noise is considered to be purely random, while spectroscopic parameters are purely systematic.

As shown in Fig. B1, when considering statistical error sources, the main contributors are the atmospheric temperature profile for setups without a temperature fit (error values of 2.0%–2.5%) and possible misalignments of the FTIR's solar tracker, given by the LOS, for all setups (error values of up to 1.0% at larger O<sub>3</sub> concentrations). By contrast, O<sub>3</sub> TCs are almost insensitive to errors due to ILS uncertainties and measurement noise (error values are smaller than 0.1%–0.2%). For the setups with a simultaneous temperature retrieval, this fit generates a significant cross-interference with the ILS function, leading to an increment of the ILS error contribution (Schneider and Hase, 2008; García et al., 2012), but also with the measurement noise and smoothing error (es-



**Figure B1.** Estimated statistical (ST) and systematic (SY) errors (%) for  $O_3$  TCs retrieved from all setups for different error sources (a, atmospheric temperature profile, b, ILS function, c, LOS, d, spectroscopic parameters, e, measurement noise, and f, smoothing error) as a function of  $O_3$  slant column (DU) for the measurements of Fig. 3.

**Table B1.** Error sources and assumed values used in the theoretical uncertainty estimation. The last column shows the contributions of statistical (ST) and systematic (SY) sources to the total error. chann.: channelling; 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

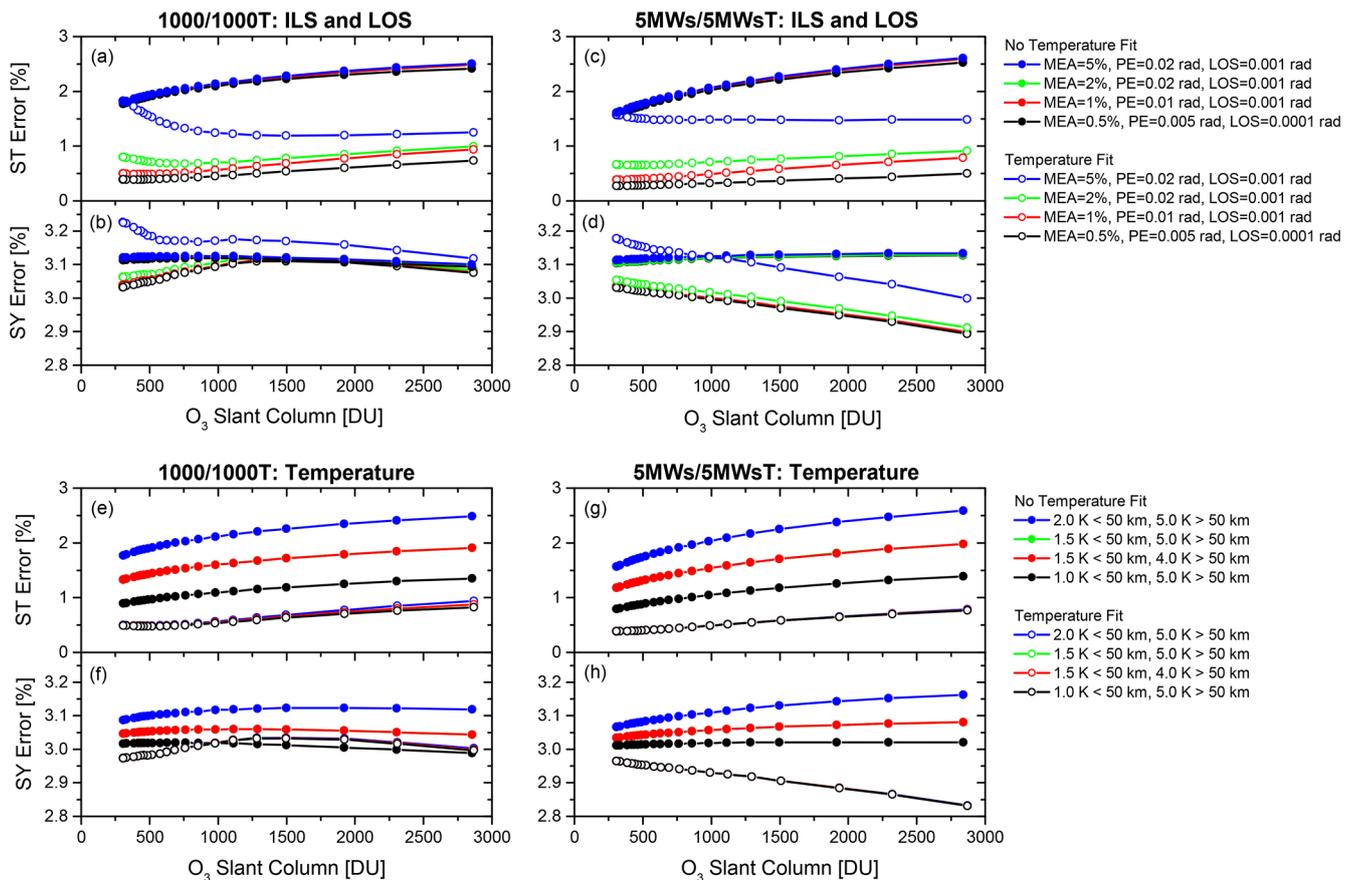
pecially for the 1000T setup). However, in return, the temperature error contribution is nearly eliminated, leading to a considerable improvement of  $\sim 1\%$  in the total ST budget. It is worth highlighting that the negative cross-interference between the temperature retrieval and SE is significant for  $O_3$  SCs beyond  $\sim 500$  DU due to the loss of FTIR vertical sensitivity as  $O_3$  SCs increase (this threshold encompasses less than 20% of the  $O_3$  observations at IZO). For the typical  $O_3$  concentrations observed at IZO, the SE is improved by the temperature fit and is smaller for the 4MWsT/5MWsT setups. This result is corroborated by a comparison to coincident Brewer observations, as shown in Table 2.

Note that in the current work, the measurement noise depends on the quality of the fitted spectra (Hase et al., 2004).

Therefore, large values of the measurement noise error are observed where the fit residuals are slightly larger (especially in the broad spectral region, see Fig. B1).

Table B1 lists representative uncertainty values of the IFS 120/HR instrument in the period 2005–2008. However, the error sources associated with instrument status (i.e. ILS function, solar pointing, and measurement noise) can differ depending on the FTIR spectrometer quality and stability. In order to account for this fact, Fig. B2 summarises the effects of the different sets of error values on  $O_3$  TCs for the measurement day of Fig. B1. Note that this figure only includes the error estimations for different ILS and LOS configurations (keeping the temperature profile and measurement errors fixed) and for the 1000/1000T and 5MWs/5MWsT setups (the 4MWs/4MWsT error estimations are quite similar to the 5MWs/5MWsT ones, so they have been omitted for simplicity).

The effect of 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, statistical TPE values range from  $\sim 0.5\%$  for uncertainties of 0.5% in the MEA and 0.005 rad in the PE (representative of the well-aligned IFS 120/5HR instrument for the period 2008–2018) up to  $\sim 1.5\%$  for an MEA error of 5% and a PE error of 0.02 rad (representative of the unstable IFS 120M spectrometer). These estimated uncertainties reproduce the changes observed in FTIR  $O_3$  quality for the different periods well when compared to Brewer observations (see Table 2). The cross-interference between the temperature fit and other error sources is also evident for systematic contributions, es-



**Figure B2.** Estimated statistical (ST) and systematic (SY) TPE (%) for O<sub>3</sub> TCs retrieved from the setups 1000/1000T (a, b, e, f) and 5MWs/5MWsT (c, d, g, h) for different error configurations of the ILS function, LOS, and atmospheric temperature profile as a function of O<sub>3</sub> slant column (DU) for the measurements of Fig. 3.

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

**Data availability.** The FTIR and Brewer data are available by request from the corresponding authors, while the ECC ozonesonde data are available from the NDACC archive (<https://www-air.larc.nasa.gov/missions/ndacc/data.html?station=izana/ames/o3sonde/>, Torres et al., 2021).

**Author contributions.** OEG and EsS designed and wrote the structure and methodology of the current paper and computed the calculations required. MS, FH, and TB participated in the retrieval analysis. FH is the author of the LINEFIT and PROFFIT codes. OEG, MS, EsS, and EIS took the routine FTIR measurements and performed the maintenance and quality control of the FTIR instruments. AR, SFL, and VC are responsible for the maintenance and quality control of the Brewer spectrometers, as well as for estimating the NDACC Brewer ozone observations. CT and NP are in

charge of the ozonesonde programme at Izaña Observatory. Finally, all authors discussed the results and contributed to the final paper.

**Competing interests.** The contact author has declared that neither they nor their co-authors have any competing interests.

**Disclaimer.** Publisher's note: Copernicus Publications remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

**Acknowledgements.** The Izaña FTIR station has been supported by the German Bundesministerium für Wirtschaft und Energie (BMWi), the Helmholtz Association, the Karlsruhe Institute of Technology (KIT) and the State Meteorological Agency of Spain. The authors also acknowledge the support of the KIT and Izaña staff for maintaining the instrumentation, thus ensuring the quality of the data.

*Financial support.* The Izaña FTIR station has been supported by the German Bundesministerium für Wirtschaft und Energie (BMWi) via DLRunter grants 50EE1711A and by the Helmholtz Association via the research programme ATMO. In addition, this research research has been supported by the European Research Council under FP7/(2007–2013)/ERC grant agreement no. 256961 (project MUSICA), by the Deutsche Forschungsgemeinschaft for the project MOTIV (Geschäftszeichen SCHN 1126/2-1), by the Ministerio de Economía y Competitividad of Spain through the projects CGL2012-37505 (project NOVIA) and CGL2016-80688-P (project INMENSE), and by EUMETSAT under its fellowship programme (project VALIASI).

*Review statement.* This paper was edited by Christof Janssen and reviewed by Corinne Vigouroux and two anonymous referees.

## 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, *Q. J. Roy. Meteor. Soc.*, 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, *J. Geophys. Res.*, 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, Halkidiki, Greece, 3–7 September 1984*, vol. 1, 606–616, ISBN 902771942X, 1985.
- 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, *Atmos. Chem. Phys.*, 13, 1973–1998, <https://doi.org/10.5194/acp-13-1973-2013>, 2013.
- 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., China, 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*, edited by: 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, <https://doi.org/10.31978/666-20-010-X>, 2020.
- Drouin, B. J., Crawford, T. J., and Yu, S.: Validation of ozone intensities at 10  $\mu\text{m}$  with THz spectrometry, *J. Quant. Spectrosc. Ra.*, 203, 282–292, <https://doi.org/10.1016/j.jqsrt.2017.06.035>, 2017.
- Duflo, 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, *Atmos. Meas. Tech.*, 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, *Atmos. Chem. Phys.*, 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, *Atmos. Meas. Tech.*, 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 ( $> 3000 \text{ cm}^{-1}$ ), *Atmos. Meas. Tech.*, 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, Á. J., Hultberg, T., Redondas, A., Barthlott, S., Wiegeler, A., González, Y., and Sanromá, E.: Consistency and quality assessment of the Metop-A/IASI and Metop-B/IASI operational trace gas products ( $\text{O}_3$ ,  $\text{CO}$ ,  $\text{N}_2\text{O}$ ,  $\text{CH}_4$ , and  $\text{CO}_2$ ) in the subtropical North Atlantic, *Atmos. Meas. Tech.*, 9, 2315–2333, <https://doi.org/10.5194/amt-9-2315-2016>, 2016.
- García, O. E., Schneider, M., Sepúlveda, E., Hase, F., Blumenstock, T., Cuevas, E., Ramos, R., Gross, J., Barthlott, S., Röhling, A. N., Sanromá, E., González, Y., Gómez-Peláez, Á. J., Navarro-Comas, M., Puñtedura, O., Yela, M., Redondas, A., Carreño, V., León-Luís, S. F., Reyes, E., García, R. D., Rivas, P. P., Romero-Campos, P. M., Torres, C., Prats, N., Hernández, M., and López, C.: Twenty years of ground-based NDACC FTIR spectrometry at Izaña Observatory – overview and long-term comparison to other techniques, *Atmos. Chem. Phys.*, 21, 15519–15554, <https://doi.org/10.5194/acp-21-15519-2021>, 2021.
- Gaudel, A., Cooper, O., Ancellet, G., Barret, B., Boynard, A., Burrows, J., Clerbaux, C., Coheur, P.-F., Cuesta, J. and Cuevas, E., Doniki, S., Dufour, G., Ebojic, F., Foret, G., García, O., Granados Muños, M., Hannigan, J., Hase, F., Huang, G., Hassler, B., 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., Osman, M., Palm, M., Petetin, H., Petropavlovskikh, I., Querel, R., Rappoe,

- N., Rozanov, A., Schultz, M., Schwab, J., Siddans, R., Smale, D., Steinbacher, M., Tanimoto, H., Tarasick, D., Thouret, V., Thompson, A., Trickl, T., Weatherhead, E., Wespes, C., Worden, H., 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.
- GCOS: The Status of the Global Climate Observing System 2021: The GCOS Status Report (GCOS-240), Tech. rep., World Meteorological Organization, Geneva, Switzerland, 2021.
- Gordon, I., Rothman, L., Hargreaves, R., Hashemi, R., Karlovets, E., Skinner, F., Conway, E., Hill, C., Kochanov, R., Tan, Y., Wcislo, P., Finenko, A., Nelson, K., Bernath, P., Birk, M., Boudon, V., Campargue, A., Chance, K., Coustenis, A., Drouin, B., Flaud, J., Gamache, R., Hodges, J., Jacquemart, D., Mlawer, E., Nikitin, A., Perevalov, V., Rotger, M., Tennyson, J., Toon, G., Tran, H., Tyuterev, V., Adkins, E., Baker, A., Barbe, A., Canè, E., Császár, A., Dudaryonok, A., Egorov, O., Fleisher, A., Fleurbaey, H., Foltynowicz, A., Furtenbacher, T., Harrison, J., Hartmann, J., Horneman, V., Huang, X., Karman, T., Karns, J., Kassi, S., Kleiner, I., Kofman, V., Kwabia-Tchana, F., Lavrentieva, N., Lee, T., Long, D., Lukashevskaya, A., Lyulin, O., Makhnev, V., Matt, W., Massie, S., Melosso, M., Mikhailenko, S., Mondelain, D., Müller, H., Naumenko, O., Perrin, A., Polyansky, O., Raddaoui, E., Raston, P., Reed, Z., Rey, M., Richard, C., Tóbiás, R., Sadiek, I., Schwenke, D., Starikova, E., Sung, K., Tamassia, F., Tashkun, S., Vander Auwera, J., Vasilenko, I., Vidasin, A., Villanueva, G., Vispoel, B., Wagner, G., Yachmenev, A., and Yurchenko, S.: The HITRAN2020 molecular spectroscopic database, *J. Quant. Spectrosc. Ra.*, 277, 107949, <https://doi.org/10.1016/j.jqsrt.2021.107949>, 2022.
- 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  $\mu\text{m}$ ) and Ultraviolet (300–350 nm) Spectral Regions, *J. Phys. Chem.*, 114, 10045–10048, <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/> (last access: 31 March 2022), 2017.
- Gröbner, J., Schill, H., Egli, L., and Stübi, R.: Consistency of total column ozone measurements between the Brewer and Dobson spectroradiometers of the LKO Arosa and PMOD/WRC Davos, *Atmos. Meas. Tech.*, 14, 3319–3331, <https://doi.org/10.5194/amt-14-3319-2021>, 2021.
- Hase, F.: Error Estimation in PROFFIT revisited, in: NDSC IRWG Meeting, Tenerife, Spain, 1–3 May 2007, NDACC/IRWG, <https://www2.acom.ucar.edu/irwg> (last access: 31 March 2022), 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, *Atmos. Meas. Tech.*, 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, *J. Quant. Spectrosc. Ra.*, 87, 25–52, <https://doi.org/10.1016/j.jqsrt.2003.12.008>, 2004.
- IRWG: Uniform Retrieval Parameter Summary, NDACC/IRWG, [http://www.acom.ucar.edu/irwg/IRWG\\_Uniform\\_RP\\_Summary-3.pdf](http://www.acom.ucar.edu/irwg/IRWG_Uniform_RP_Summary-3.pdf) (last access: 31 March 2022), 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, *Int. J. Climatol.*, 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, *Atmos. Meas. Tech.*, 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, *J. Quant. Spectrosc. Ra.*, 111, 569–585, <https://doi.org/10.1016/j.jqsrt.2009.10.013>, 2010.
- Piquet-Varrault, B., Orphal, J., Doussin, J. F., Carlier, P., and Flaud, J. M.: Laboratory Intercomparison of the Ozone Absorption Coefficients in the Mid-infrared (10  $\mu\text{m}$ ) and Ultraviolet (300–350 nm) Spectral Regions, *J. Phys. Chem.*, 109, 1008–1014, <https://doi.org/10.1021/jp0405411>, 2005.
- 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, *Atmos. Chem. Phys.*, 14, 1635–1648, <https://doi.org/10.5194/acp-14-1635-2014>, 2014.
- Rimmer, J. S., Redondas, A., and Karppinen, T.: EuBrewNet – A European Brewer network (COST Action ES1207), an overview, *Atmos. Chem. Phys.*, 18, 10347–10353, <https://doi.org/10.5194/acp-18-10347-2018>, 2018.
- Rodgers, C.: Inverse Methods for Atmospheric Sounding: Theory and Praxis, World Scientific Publishing Co., Singapore, ISBN 981-02-2740-X, 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., 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, *J. Quant. Spectrosc. Rad.*, 110, 533–572, <https://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, *Atmos. Chem. Phys.*, 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<sub>3</sub> monitoring techniques, *Atmos. Chem. Phys.*, 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<sub>3</sub> profiles measured by a state-of-the-art ground-based FTIR observing system, *Atmos. Chem. Phys.*, 8, 5579–5588, <https://doi.org/10.5194/acp-8-5579-2008>, 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., Wiegeler, 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, *Atmos. Meas. Tech.*, 5, 3007–3027, <https://doi.org/10.5194/amt-5-3007-2012>, 2012.
- Steinbrecht, W., Froidevaux, L., Fuller, R., Wang, R., Anderson, J., Roth, C., Bourassa, A., Degenstein, D., Damadeo, R., Zawodny, J., Frith, S., McPeters, R., Bhartia, P., Wild, J., Long, C., Davis, S., Rosenlof, K., Sofieva, V., Walker, K., Rahpoe, N., Rozanov, A., Weber, M., Laeng, A., von Clarmann, T., Stiller, G., Kramarova, N., Godin-Beekmann, S., Leblanc, T., Querel, R., Swart, D., Boyd, I., Hocke, K., Kämpfer, N., Maillard Barras, E., Moreira, L., Nedoluha, G., Vigouroux, C., Blumenstock, T., Schneider, M., García, O., Jones, N., Mahieu, E., Smale, D., Kotkamp, M., Robinson, J., Petropavlovskikh, I., Harris, N., Hassler, B., Hubert, D., and Tummon, F.: An update on ozone profile trends for the period 2000 to 2016, *Atmos. Chem. Phys.*, 17, 10675–10690, <https://doi.org/10.5194/acp-17-10675-2017>, 2017.
- Torres, C., Prats, N., Redondas, A., and Cuevas, E.: ECC Ozone sondes data from Izaña (Spain), Network for the Detection of Atmospheric Composition Change [data set], <https://www-air.larc.nasa.gov/missions/ndacc/data.html?station=izana/ames/o3sonde/> (last access: 31 March 2022), 2021.
- Tyuterev, V. G., Barbe, A., Jacquemart, D., Janssen, C., Mikhailenko, S. N., and Starikova, E. N.: Ab initio predictions and laboratory validation for consistent ozone intensities in the MW, 10 and 5 μm ranges, *J. Chem. Phys.*, 150, 184303, <https://doi.org/10.1063/1.5089134>, 2019.
- 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<sub>3</sub> total column with data from two different IASI algorithms and from OMI and GOME-2 satellite instruments, *Atmos. Meas. Tech.*, 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., Velasco, 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, *Atmos. Chem. Phys.*, 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, *Atmos. Chem. Phys.*, 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., DeLuise, 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, *J. Geophys. Res.*, 105, 22201–22210, <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., edited by: Smit, H. G. J. and ASOPOS panel, 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, 588 pp., ISBN 978-1-7329317-1-8, 2018.