HAL
open science

Validation of ACE-FTS satellite data in the upper troposphere/lower stratosphere (UTLS) using non-coincident measurements

M. I. Hegglin, C. D. Boone, G. L. Manney, T. G. Shepherd, K. A. Walker, P. F. Bernath, W. H. Daffer, P. Hoor, C. Schiller

To cite this version:

M. I. Hegglin, C. D. Boone, G. L. Manney, T. G. Shepherd, K. A. Walker, et al.. Validation of ACE-FTS satellite data in the upper troposphere/lower stratosphere (UTLS) using non-coincident measurements. Atmospheric Chemistry and Physics, 2008, 8 (6), pp.1483-1499. hal-00296485

HAL Id: hal-00296485

https://hal.science/hal-00296485

Submitted on 18 Jun 2008

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Atmos. Chem. Phys., 8, 1483–1499, 2008
www.atmos-chem-phys.net/8/1483/2008/
© Author(s) 2008. This work is distributed under
the Creative Commons Attribution 3.0 License.

Atmospheric Chemistry and Physics

Validation of ACE-FTS satellite data in the upper troposphere/lower stratosphere (UTLS) using non-coincident measurements

M. I. Hegglin[1], C. D. Boone[2], G. L. Manney[3,4], T. G. Shepherd[1], K. A. Walker[1,2], P. F. Bernath[2,5], W. H. Daffer[6], P. Hoor[7], and C. Schiller[8]

[1]Department of Physics, University of Toronto, Toronto, Canada
[2]Department of Chemistry, University of Waterloo, Waterloo, Canada
[3]Jet Propulsion Laboratory, California Institute of Technology, Pasadena, California, USA
[4]New Mexico Institute of Mining and Technology, Socorro, New Mexico, USA
[5]Department of Chemistry, University of York, York, UK
[6]Columbus Technologies Inc., Pasadena, California, USA
[7]Max Planck Institute for Chemistry, Air Chemistry, Mainz, Germany
[8]Institute for Chemistry and Dynamics of the Geosphere 1: Stratosphere, Research Centre Jülich GmbH, Jülich, Germany

Received: 27 August 2007 – Published in Atmos. Chem. Phys. Discuss.: 26 September 2007
Revised: 3 January 2008 – Accepted: 12 February 2008 – Published: 13 March 2008

Abstract. CO, O_3, and H_2O data in the upper troposphere/lower stratosphere (UTLS) measured by the Atmospheric Chemistry Experiment Fourier Transform Spectrometer (ACE-FTS) on Canada's SCISAT-1 satellite are validated using aircraft and ozonesonde measurements. In the UTLS, validation of chemical trace gas measurements is a challenging task due to small-scale variability in the tracer fields, strong gradients of the tracers across the tropopause, and scarcity of measurements suitable for validation purposes. Validation based on coincidences therefore suffers from geophysical noise. Two alternative methods for the validation of satellite data are introduced, which avoid the usual need for coincident measurements: tracer-tracer correlations, and vertical tracer profiles relative to tropopause height. Both are increasingly being used for model validation as they strongly suppress geophysical variability and thereby provide an "instantaneous climatology". This allows comparison of measurements between non-coincident data sets which yields information about the precision and a statistically meaningful error-assessment of the ACE-FTS satellite data in the UTLS. By defining a trade-off factor, we show that the measurement errors can be reduced by including more measurements obtained over a wider longitude range into the comparison, despite the increased geophysical variability. Applying the methods then yields the following upper bounds to the relative differences in the mean found between the ACE-FTS and SPURT aircraft measurements in the upper troposphere (UT) and lower stratosphere (LS), respectively: for CO ±9% and ±12%, for H_2O ±30% and ±18%, and for O_3 ±25% and ±19%. The relative differences for O_3 can be narrowed down by using a larger dataset obtained from ozonesondes, yielding a high bias in the ACE-FTS measurements of 18% in the UT and relative differences of ±8% for measurements in the LS. When taking into account the smearing effect of the vertically limited spacing between measurements of the ACE-FTS instrument, the relative differences decrease by 5–15% around the tropopause, suggesting a vertical resolution of the ACE-FTS in the UTLS of around 1 km. The ACE-FTS hence offers unprecedented precision and vertical resolution for a satellite instrument, which will allow a new global perspective on UTLS tracer distributions.

Correspondence to: M. I. Hegglin
(michaela@atmosp.physics.utoronto.ca)

1 Introduction

The upper troposphere/lower stratosphere (UTLS) has recently attracted major research interest in atmospheric science due to its key role in chemistry-climate coupling. In order to characterize UTLS tracer distributions and to detect future changes, tracer measurements with global coverage are needed. Satellites are the only means by which this task can be accomplished, however, their capability to measure accurately in the UTLS is limited, and the validation of the measurements difficult.

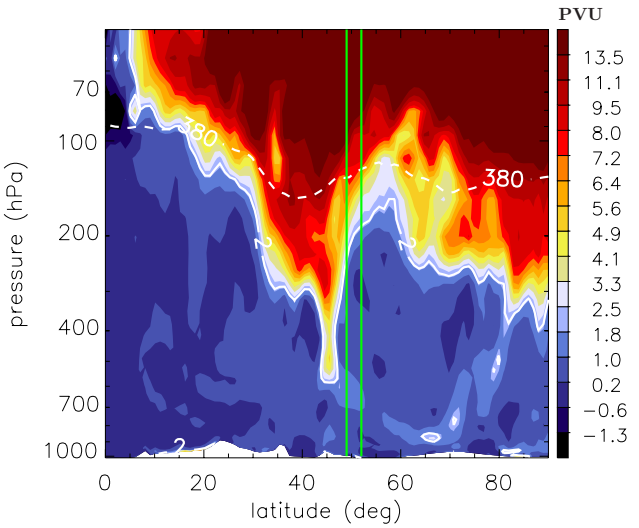

Fig. 1. Vertical cross section of ECMWF potential vorticity (colour coded) from equator to North Pole at 6° W on 10 November 2001 at 06:00 UTC. White thick line shows the dynamical tropopause (2 PVU, with 1 PVU=1 potential vorticity unit=10^{-6} $m^2 s^{-1} K kg^{-1}$), white dashed line the 380 K isentrope. Green lines indicate the location of two independent profiles that fulfill the spatial coincidence criteria of being taken no further apart than 500 km.

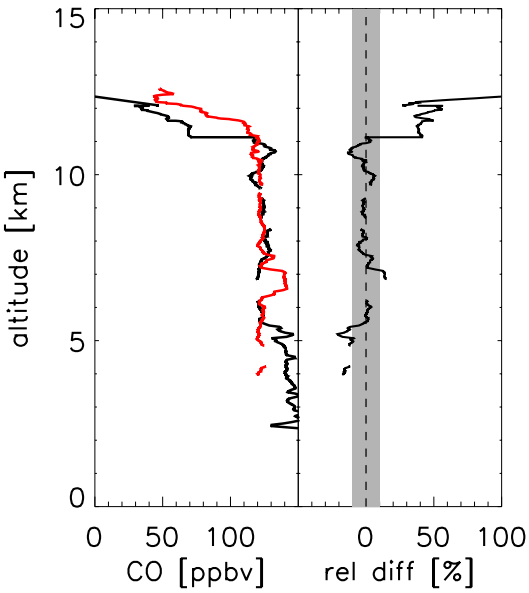

Fig. 2. Left panel: Vertical profiles of CO mixing ratios taken during ascent (black) and descent (red) of the SPURT aircraft on 10 November 2001 from and to the airport Hohn (Northern Germany). The profiles were taken approximately 10 h apart. Right panel: Relative difference of the measured profiles. Gray shading indicates an error range of ±10%.

These limitations arise from the dynamical and chemical structure of the UTLS. Dynamical variability in the tropopause region induced by Rossby wave activity is high, and length scales of the associated features in the tracer fields are small – less than 1 km in the vertical, and 100 km in the horizontal. Tracer mixing ratios also exhibit a strong gradient across the tropopause because it acts as a transport barrier (Pan et al., 2004; Hoor et al., 2004; Hegglin et al., 2006 and references therein). Ultimately, the remote sensing technique used determines the achievable vertical and horizontal resolution of the measurements and hence the capability of the instrument to resolve the given geophysical small-scale variability. The Atmospheric Chemistry Experiment Fourier Transform Spectrometer (ACE-FTS) on Canada's SCISAT-1 satellite (Bernath et al., 2005) has provided accurate measurements of numerous chemical species throughout the stratosphere and into the UT since February 2004, with a vertical spacing between measurements of less than 1 km at the lowest retrieval altitudes. The measurements suggest a high potential for studies related to the UTLS.

In the stratosphere and mesosphere, the ACE-FTS satellite measurements are being validated by comparison to balloon-borne or independent satellite data which are approximately coincident in time and space (see other publications in this special issue on ACE validation). Coincidence criteria are defined in various ways, but typically the measurements have to be taken within several hours and at locations no further apart than around 500 km. However, in the case of occultation sounders, it can be difficult to find a sufficient number of coincident measurements for a statistically meaningful validation (Walker et al., 2005). Furthermore, this validation method does not account for differences in the tracer mixing ratios which are produced by the geophysical variability found within the defined time and length scales.

The use of coincident measurements is especially an issue in the UTLS, where geophysical variability is large and a strong gradient in chemical tracers is found across the tropopause. Figure 1 provides a Northern Hemisphere cross section of ECMWF potential vorticity at 6° W on 10 November 2001 at 06:00 UTC. The dynamical tropopause (white line) shows strong undulations and even profiles taken no further apart than 500 km exhibit large differences in tropopause height (green lines). Figure 2 shows two vertical profiles of CO mixing ratios measured in-situ by an aircraft instrument during the same dynamical situation and taken within 8 h and a distance of 400 km. As can be seen, the apparent error derived for the measurements in the tropopause region is as large as 50%. However, this is due to the geophysical variability and not a measurement error (the profiles are taken by the same instrument). Further complicating a statistically meaningful comparison is the sharp tracer gradient at the tropopause, which causes sampling errors to be non-normally distributed.

On the other hand, sparse data sets are increasingly being used for the validation of chemistry transport and chemistry

climate models, through diagnostics that provide "instantaneous climatologies" and reveal characteristic features in tracer distributions (Douglass et al., 2003; Sankey and Shepherd, 2003; Hegglin and Shepherd, 2007; Strahan et al., 2007). In the UTLS, such characteristic features are found when moving from geometric altitude coordinates into tracer-tracer space or tropopause coordinates. These transformations strongly reduce geophysical variability and help in assessing the quality and vertical information content of the model data. The same approach can therefore be considered for the validation of new satellite data.

In this study, we investigate the use of tracer-tracer correlations and vertical tracer profiles relative to the tropopause height for the validation of the ACE-FTS CO, O_3, and H_2O measurements in the UTLS. Section 2 provides the description of the data sets used. In Sect. 3, the new validation methods are introduced, and their strengths and weaknesses discussed. We assess whether spatial and temporal variations in the "climatologies" are acceptable for identifying various types of errors. The methods are then used in Sect. 4 to validate the ACE-FTS measurements using aircraft and ozonesonde data. A summary is given in Sect. 5.

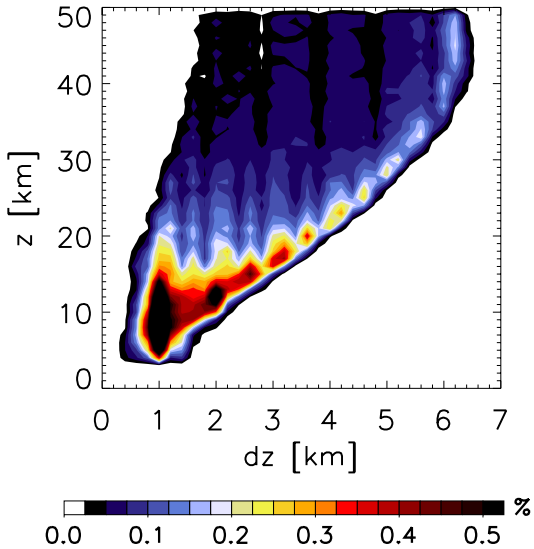

Fig. 3. Probability density function of the vertical spacing between the ACE-FTS measurements (dz) as a function of retrieval altitude. The vertical spacing in the tropopause region (between approximately 8 and 14 km) exhibits values that can be much lower than 1 km.

2 Data description

2.1 ACE-FTS satellite data

The Atmospheric Chemistry Experiment Fourier Transform Spectrometer (ACE-FTS) is the primary instrument on SCISAT-1, a Canadian-led satellite mission for remote sensing of the Earth's atmosphere. ACE-FTS features high resolution (0.02 cm^{-1}) and broad spectral coverage in the infrared (750 to 4400 cm^{-1}). The instrument operates almost exclusively in solar occultation mode (Bernath et al., 2005). The SCISAT-1 satellite was launched into low Earth circular orbit (650 km) with high inclination (74°) on 12 August 2003. In solar occultation mode, this orbit provides seasonally varying coverage of the globe, with an emphasis on mid-latitudes and polar regions. Up to 30 occultation events (sunrises or sunsets viewed by the orbiting satellite) occur per calendar day. Science operations for the ACE-FTS began in February 2004. Retrievals for the ACE-FTS employ a non-linear least squares global-fit approach (Boone et al., 2005).

The instrument has a 1.25 mrad input aperture, which subtends an altitude range of 3–4 km at the tangent point (the point of closest approach to the Earth for a solar ray measured by the instrument). However the ACE-FTS instrument collects measurements every 2 s, which can lead to significant oversampling. The altitude spacing between measurements varies over the course of the year, governed primarily by the beta angle (the angle between the satellite orbit plane and the Earth-Sun vector) corresponding to the occultation. The rate of change of the tangent height of each measurement within an occultation decreases with increasing beta angle, leading to higher vertical sampling for larger beta angles. At low altitudes, refraction effects also impact the measurement spacing. ACE-FTS results are provided on two altitude grids, a 1-km grid common to all occultations and a "retrieval grid" that varies from occultation to occultation. The retrieval grid contains values at the measurement altitudes, unless there are multiple measurements within a layer on the 1-km grid, in which case it provides a single value at the center of the 1-km grid layer. For this study, we use the measurements on the retrieval grid. Figure 3 shows the probability density function of the vertical spacing between altitudes on the retrieval grid as a function of height for all measurements between 5 and 50 km. The altitude spacing in the UTLS varies from about 3 km to less than 1 km. Part of the purpose of this paper is to assess how much real vertical resolution there is in the retrieval grid.

In this study we focus on the validation of the ACE-FTS version 2.2 H_2O, CO, as well as the "version 2.2 O_3 update" results in the UTLS between February 2004 and January 2007. The UTLS comprises the tropopause region, and lies between approximately 5 and 15 km altitude or 500 and 100 hPa. The validation of the stratospheric and mesospheric ACE-FTS version 2.2 data (with ozone updates) is published along with the paper presented here in the special issue on validation of ACE (for CO: Clerbaux et al. (2007), for O_3: Dupuy et al. (2008), for H_2O: Carleer et al. (2008)). Earlier comparisons of the version 1.0 ACE-FTS O_3 data with GOMOS, POAM III, and SAGE III satellite data (Fussen et al., 2005; Walker et al., 2005) yielded good agreement between the data sets with differences mostly lower than 10%

between 15 and 45 km. With HALOE, an agreement of ±5% between 15 and 35 km was found (McHugh et al., 2005). In the same study, the ACE-FTS H_2O data indicated a high bias of around 20% at altitudes below 20 km. A first validation of the version 1.0 CO by Jin et al. (2005) using Odin satellite data yielded excellent agreement, and a comprehensive analyses of these data was provided by Clerbaux et al. (2005). ACE-FTS version 2.1 level data were compared to the Aura-MLS version 1.5 (Froidevaux et al., 2006), and Aura-MLS version 2.2 satellite data (Lambert et al., 2007[1]; Froidevaux et al., 2008; Pumphrey et al., 2007[2]). The MLS/ACE-FTS differences for O_3 in the lower stratosphere are within ±5%, but increase with altitude. For H_2O, the instruments show good agreement, within ±5%, between 68–0.004 hPa. At the lowest levels considered in this study (around 100 hPa) the differences increase up to 30%. Differences in the CO measurements are around 50% in the stratosphere, and 25% in the mesosphere.

2.2 SPURT aircraft data

As a reference data set, we use in-situ high-resolution and high-precision CO, O_3, and H_2O measurements from the SPURT (German acronym for "trace gas transport in the tropopause region") aircraft campaign. The flights were carried out seasonally between November 2001 and July 2003 in the Northern Hemisphere over Europe and cover a latitude range between 30° N and 80° N. For each season, approximately 32 flight hours or 24 000 data points were obtained. An overview of the campaign can be found in Engel et al. (2006). Detailed descriptions of the CO, O_3, and H_2O measurements can be found in Hoor et al. (2004), Hegglin et al. (2006), and Krebsbach et al. (2006), respectively.

2.3 WOUDC ozonesonde data

In order to corroborate the results obtained by the validation using the spatially limited aircraft measurements we extend our evaluation to the WOUDC ozonesonde data set. The data set is available at the WOUDC Data Web page (http://www.woudc.org/data_e.html) and includes over 300 stations worldwide. We use temperature and ozone profiles with a vertical resolution of 100-150 m obtained during the years 2004–2007 and in a latitude range between 40° N and 60° N. The accuracy of ozonesonde observations is estimated to be ±5% (e.g. SPARC, 1998).

[1]Lambert, A., Read, W. G., Livesey, N. J., et al.: Validation of the Aura Microwave Limb Sounder middle atmosphere water vapor and nitrous oxide measurements, J. Geophys. Res., submitted, 2007.

[2]Pumphrey, H. C., Filipiak, M. J., Livesey, N. J., et al.: Validation of middle-atmosphere carbon monoxide retrievals from MLS on Aura, J. Geophys. Res., submitted, 2007.

2.4 Derived meteorological products

For each measurement of the different data sets, we derived the thermal tropopause height according to the WMO definition, i.e. the lowest level at which the lapse rate drops to $2\,\mathrm{K\,km^{-1}}$ or less, and the average lapse rate between this level and all higher levels within 2 km does not exceed $2\,\mathrm{K\,km^{-1}}$ (WMO, 1957). The tropopauses for the ozonesonde data were calculated using simultaneously measured temperature profiles, and those for the ACE-FTS and the SPURT data using the Goddard Earth Observing System Model, Version 4 (GEOS-4), by interpolation of the model fields onto the exact measurement location in time and space and by applying the Reichler et al. (2003) algorithm. The GEOS-4 analyses are described by Bloom et al. (2005). The GEOS-4 data used here are provided on 55 hybrid (σ/pressure) model levels from the surface to 0.01 hPa. The horizontal grid is 1.0° latitude by 1.25° longitude. Six-hourly average fields are provided centered at 00:00, 06:00, 12:00 and 18:00 UTC. Besides the standard meteorological variables, GEOS-4 products include an extensive set of fields from the model and assimilation system, including PV calculated internally in the model. Further information on the derived meteorological products can be found in Manney et al. (2007).

2.5 Canadian Middle Atmosphere Model data

The Canadian Middle Atmosphere Model (CMAM) is used to investigate the geophysical variability in the tracer distributions on different time and spatial scales, which potentially influences the reliability of the validation methods introduced here. CMAM is an extension of the Canadian Centre for Climate Modelling and Analysis (CCCma) spectral General Circulation Model into the lower thermosphere up to 0.0006 hPa (around 100 km). CMAM includes a comprehensive representation of the relevant physical and chemical processes in a fully interactive mode (Beagley et al., 1997; de Grandpré et al., 2000). Data presented here correspond to the 3-day output of instantaneous fields in the years 2000–2010 of a 150-year transient run from 1950 with CMAM version 8, using model-generated sea-surface temperatures and background aerosol forcing. Model fields are calculated on a linear Gaussian transform grid with 32×64 grid points in the horizontal, corresponding to a resolution of around 6°×6°, and 71 vertical levels. The vertical resolution in the tropopause region is around 900 m, increasing to around 2 km in the upper stratosphere. A detailed comparison of CMAM version 8 (run over 1960–2004 with observed sea surface temperatures and volcanic aerosol forcing) with observations is provided as part of the model intercomparison of Eyring et al. (2006).

3 New validation methods

The underlying idea of the validation methods introduced here is based on the widely accepted use of characteristic features found in tracer climatologies to test the simulation of tracer distributions in chemistry climate or chemistry transport models (Douglass et al., 2003; Sankey and Shepherd, 2003; Hegglin and Shepherd, 2007; Strahan et al., 2007). These climatologies are often obtained from spatially and temporally inhomogeneous tracer observations, yet are considered to represent the full atmosphere. In the UTLS, characteristic features in tracer distributions are found when moving from geometric altitude coordinates into tracer-tracer space or tropopause coordinates, as explained in more detail below. Our hypothesis is that new observations, such as the ACE-FTS data set, can be considered accurate and precise if shown to be capable of resolving the characteristic features found in such climatologies. Since this approach allows one to include all measurements, not just the coincident ones, one can get much better statistics. When using tracer-tracer correlations in the lower stratosphere, for example, model errors of the order of 10% can be detected (Hegglin and Shepherd, 2007).

3.1 Tracer-tracer correlations

Sufficiently long-lived species exhibit compact correlations (Plumb and Ko, 1992), which reduce day-to-day variations and provide an "instantaneous climatology". Depending on the lifetimes of the tracers used, compact correlations are not necessarily linear; as in the case of the O_3-CO and O_3-H_2O correlations, they can exhibit a strong curvature which can be used to identify the chemical transition between the troposphere and the stratosphere (Pan et al., 2007 and references therein). Apart from recent troposphere-to-stratosphere transport events, which produce distinct and nearly linear mixing lines (Fischer et al., 2000), the compactness of these correlations is relatively high and their shape distinct due to their strong dependence on the location of the tropopause, so that they can be used to gauge the precision of the ACE-FTS measurements.

3.2 Vertical profiles relative to tropopause height

Another method for reducing the effects of geophysical variability in UTLS tracer measurements, and for obtaining fields suitable for comparison of non-coincident measurements, is the use of tracer vertical profiles relative to the tropopause height (i.e. in tropopause coordinates) (cf. Hoor et al., 2004; Pan et al., 2004; Hegglin et al., 2006). These profiles are more compact and show less scatter than data plotted in geometric altitude (or potential temperature) coordinates. This is shown in Fig. 4a and b for ozonesonde data obtained during winter (DJF) and accumulated over the years 2004–2007. They furthermore reveal a sharp gradient between tropospheric and stratospheric tracer mixing ratios which can be used to test the vertical resolution and information content of the ACE-FTS data. A recent study by Monahan et al. (2007) used tropopause coordinates for the validation of O_3 data from the AIRS satellite using coincident measurements. They showed that tropopause coordinates allow the separation of the measurement errors between the troposphere and the stratosphere, which led to a decrease and increase in the error in lower stratospheric and upper tropospheric measurements, respectively. We show here that this validation method, when used in a climatological rather than an instantaneous way, can be applied even to non-coincident measurements, allowing a statistically meaningful error assessment for sparse data sets.

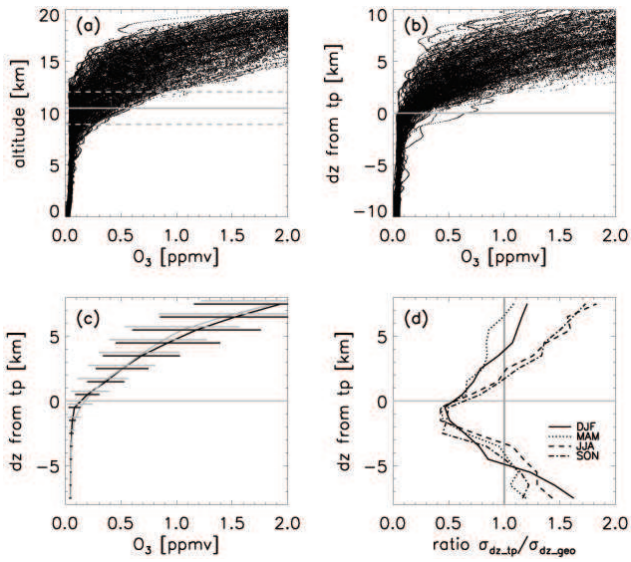

Fig. 4. Ozonesonde data obtained in Northern Hemisphere winters 2004-2007 between 40° N and 60° N. **(a)** In geometric altitude and indicating the average height of the tropopause (gray solid line) ± its standard deviation σ (dashed gray lines), **(b)** in tropopause coordinates, and **(c)** plotted as means $\pm\sigma$ with the data from (a) in gray plotted relative to the mean tropopause height (subscript *dz_geo*), and data from (b) in black (subscript *dz_tp*). **(d)** Ratio between the standard deviations of the two sample methods ($\sigma_{dz_tp}/\sigma_{dz_geo}$), now depicted for all seasons.

In order to apply the tropopause coordinates to the ACE-FTS and SPURT or ozonesonde data, we first calculate the distance from the thermal tropopause of each measurement point. We then calculate the CO, O_3, and H_2O mean mixing ratios ($\overline{x}$) and their standard deviations (σ) for each 1 km altitude bin from −6 to +6 km relative to the tropopause height. We finally derive the relative differences (Δ_{rel}) between the mean profiles of the ACE-FTS (subscript ACE) and the reference data set (subscript ref) using

$$\Delta_{\mathrm{rel}} = \frac{\overline{x}_{\mathrm{ACE}} - \overline{x}_{\mathrm{ref}}}{0.5 \times (\overline{x}_{\mathrm{ACE}} + \overline{x}_{\mathrm{ref}})}\,, \tag{1}$$

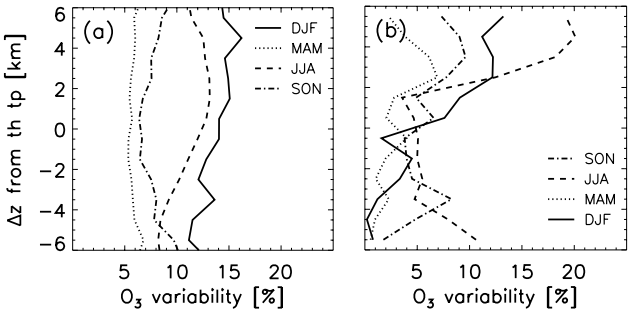

Fig. 5. Interannual variability in the seasonal means of vertical ozone profiles in tropopause coordinates for the latitude range between 40° N and 60° N and all longitudes, calculated from **(a)** ten years of CMAM data and **(b)** ozonesondes taken between 2004 and 2007.

and their relative uncertainties (δ_{rel}) calculated according to the general rules of error propagation, and using the standard errors (δ) of each mean value estimated by

$$\delta = \frac{\sigma}{\sqrt{N/b}} \,. \tag{2}$$

Here, N is the number of measurements, and b is a factor to account for the autocorrelation length of the measurements, taken to be 30, 45, and 3 for ozonesonde, SPURT aircraft, and ACE-FTS data, respectively. For the SPURT data set, b corresponds to the autocorrelation length between measurements along a flight track which is mainly determined by the length scales generated by the prevailing dynamical fields (≈350 km). For the ozonesondes and the ACE-FTS profiles, b is determined by the maximum number of measurements per profile expected to end up in the same altitude bin. Single profiles are assumed to be independent of each other. Conservative values were chosen for these parameters.

3.3 Strengths and limitations of the validation methods

The use of "instantaneous climatologies" in the form of tracer-tracer correlations or vertical profiles in tropopause coordinates allows one to compare a much larger number of measurements than using coincident measurements only. This helps reduce the standard errors of the derived mean values, and hence allows a statistically more meaningful error assessment. The downside of including more measurements taken over larger geographical regions and longer time periods, however, is that the geophysical variability increases. Furthermore the "instantaneous climatology" will itself vary, which introduces a potential source of error if the two data sets have different sampling. In the worst case, if these spatial and temporal variations are larger than the measurement errors, the real uncertainty in the measurement will never be found. Hence the questions to be addressed in our study are: how much daily geophysical variability can be eliminated? Can the elimination of this variability compensate for the inherent uncertainty when using climatologies in tropopause coordinates? How many independent measurements are needed to describe this climatology? We will answer these questions by focusing on vertical profiles in tropopause coordinates. However, the derived results will be valid as well for the tracer-tracer correlations.

3.3.1 Daily variability and climatological uncertainty

We can estimate the reduction in geophysical variability by comparing the standard deviations of the mean O_3 values calculated from the ozonesondes in each coordinate system (Fig. 4c). The data in geometric altitude is plotted relative to the mean tropopause height in order to allow direct comparison of the two coordinate systems. Moving from geometric altitude to tropopause coordinates reduces the geophysical variability by up to 50% around the tropopause (Fig. 4d). This accounts for the very large standard deviations found in the error assessment for the ACE-FTS O_3 at altitudes below 15 km (note the tropopause is found around 10–12 km) using coincident measurements (Fig. 46 in Dupuy et al., 2008), and implies that accounting for the daily geophysical variability will greatly improve the validation. It is furthermore interesting to see that the effectiveness of the tropopause coordinate system in reducing geophysical variability decreases with increasing distance from the tropopause, and reverses sign beyond a certain distance. This evaluation reveals that the influence of the tropopause in shaping UTLS tracer distributions is limited to a certain depth which is determined by the season. During winter and spring, variability is reduced up to 5 km (4 km) above (below) the tropopause, but only up to 2 km (3 km) in summer and autumn. Hence the use of tropopause coordinates is beneficial only in this altitude range.

The question about the inherent uncertainty of "instantaneous climatologies" using tropopause coordinates is more difficult to answer since it is influenced by geophysical variability on various time and length scales. This cannot be determined by observations alone since they are limited in space and time, and might be affected by sampling biases. However, we can use a model which provides information about the full atmosphere (in our case CMAM), to estimate the geophysical variability on the time and length scales chosen. In order to test the capability of CMAM to capture geophysical variability appropriately, we compare the interannual variability in the seasonal means obtained for a latitude range between 40° N and 60° N derived from ten years of CMAM data and from 3 years of ozonesonde data (Fig. 5). The evaluation shows that the variability simulated by CMAM lies around 10% for all altitude levels. The interannual variability of the seasonal means derived from the ozonesonde observations shows a similar range in the LS, and a smaller one for the UT. This evaluation provides us with some confidence that we can use CMAM to test sampling requirements.

We now determine how many independent measurements are needed to represent the mean profiles of the full atmosphere, so that the error introduced by the spatial and temporal variability inherent in the climatology is acceptable. To do this we use CMAM data obtained during a single winter season (DJF) from latitudes between 40° N and 60° N (including all longitudes). Datasets for 12 altitude levels between −6 and 6 km relative to the tropopause height (with a bin size of 1 km) are generated, which will be referred to henceforth as full sampling. From these datasets, we generate 1000 smaller ("subsampled") datasets each of which is composed of a different number of measurements ranging from 1 to 1000. Means and standard deviations are computed for each of the full sampling and the "subsampled" datasets. The relative differences of the means between full sampling and the "subsampled" datasets for each altitude level (colour coded) are shown in the top panels of Fig. 6 (O_3 in panel a, H_2O in panel b). The relative differences decrease with the number of data points included in the subsample. They are generally larger for lower (bluish colours) than for higher altitude levels (reddish colours), and larger for H_2O than for O_3. The relative differences lie on average within 10% for H_2O, but within 5% for O_3.

In order to show that the relative differences are decreasing with increasing number of included measurements following theoretical expectations, the 1000 "subsampled" datasets of each level are regenerated 20 times using randomized data, and their means and standard deviations are computed. The relative differences between the means of these 20 "subsampled" and the fully sampled datasets are shown in the lower panels of Fig. 6 (gray shadings) for selected altitude levels. The colored curves indicate the theoretically expected decrease in the relative differences, which is described by the standard error of the mean ($\sigma/\sqrt{N}$), where σ is the standard deviation of the full sampling. The decrease is dependent on the respective altitude level. The relative differences are seen to be largest just below the tropopause, and are smaller in the LS and in the troposphere. Also denoted are the number of measurements (black asterisks) needed to reduce the relative differences to ±5% for O_3, and to ±15% for H_2O. 50–100 independent measurements are needed for O_3, and 80–120 for H_2O, depending on the altitude level.

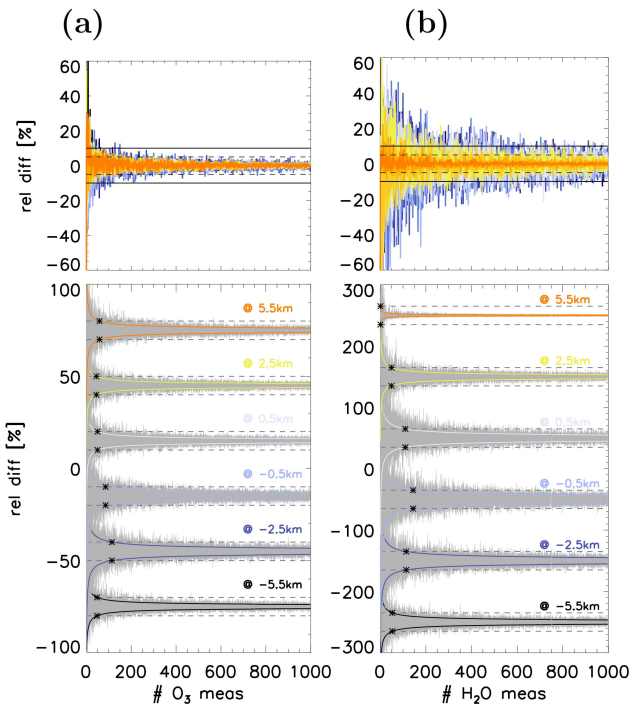

Fig. 6. Relative differences between the means of the full model field and the means derived from a specified number (N) of independent data points plotted as a function of N for **(a)** O_3, and **(b)** H_2O. Data used are restricted to the latitude range between 40° N and 60° N and DJF of one year. Upper panels: one calculation for each altitude bin between −6 and +6 km above the tropopause (1 km spacing, colour coded) with dashed/solid lines indicating ±5% and ±10% relative difference. Lower panels: 20 calculations using randomized data (in gray) for selected altitudes relative to the tropopause height (−5.5, −2.5, −0.5, 0.5, 2.5, 5.5 km, colour coded) and offset by 75, 45, 15, −15, −45, and −75% for O_3, and by 250, 150, 50, −50, −150, −250% for H_2O, respectively. Dashed lines indicate in (a) ±5%, and (b) ±15% relative difference. The coloured lines illustrate how the standard error of the mean ($\pm\sigma/\sqrt{N}$, with σ being the standard deviation of the full sampling) is decreasing with increasing number of measurements. The black asterisks indicate the number of measurements needed to decrease the standard error of the mean to ±5% and ±15% for O_3 and H_2O, respectively.

3.3.2 Trade-off factor

Given an available set of measurements, we now need to assess whether it is beneficial to move to a wider longitude range and/or add more years of data. Expressed in another way, we need to know whether the trade-off between increasing geophysical variability and adding more measurements is helping to improve our error assessment. To this end we compare the standard errors of the mean ($\sigma/\sqrt{N}$), shown in the previous section to describe the decrease in relative differences, for CMAM datasets including different longitude ranges and years. Examples of this evaluation are shown in Fig. 7. The black curves denote the standard errors of the mean computed using the σ for data obtained over one year of DJF data between 40° N and 60° N, and restricted to longitudes between 0° E and 45° E. The gray curves are for the full longitude range. The upper curves are for O_3 at 3.5 km, the lower curves for H_2O at 1.5 km above the tropopause. Red triangles indicate the number of available independent SPURT measurements (N/b), and black and gray asterisks the number of independent ACE-FTS measurements for the restricted and full longitude ranges, respectively. Note that the ozonesondes provide over 200 independent measurements for both full and restricted longitude ranges and all altitude

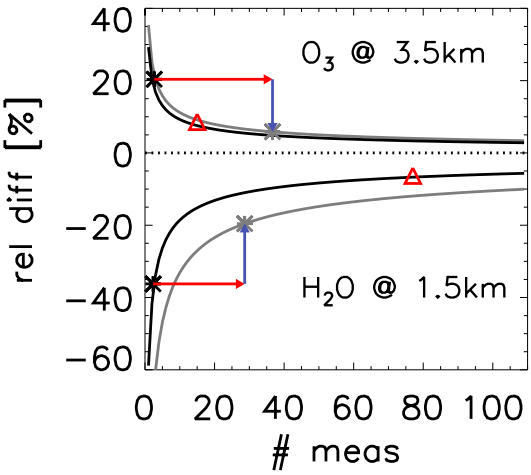

Fig. 7. Standard errors of the mean ($\sigma/\sqrt{N}$) as a function of number of measurements (N) for O_3 at 3.5 km (upper curves) and H_2O at 1.5 km above the tropopause (lower curves, sign has been flipped) and for two different CMAM datasets including DJF data from 5 years and a latitude range between 40° N and 60° N. Black lines denotes the calculation for data restricted to a longitude range of 0° E–45° E, gray lines include measurements from all longitudes. Black and gray asterisks denote the number of independent ACE-FTS measurements (N/b) for the restricted and full longitude ranges during DJF and between 2004–2007. Red triangles denote the number of independent SPURT measurements. Red arrow indicates the gain in number of measurements, blue arrow the decrease in relative difference when moving from the restricted to the full longitude range.

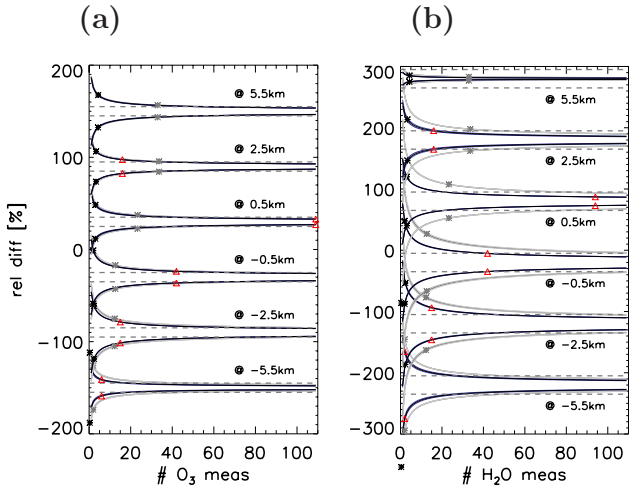

Fig. 8. Same as in previous Fig. 7, but now for all levels and for **(a)** O_3, and **(b)** H_2O. Darker lines are calculations for the CMAM dataset restricted to a latitude range of 40° N–60° N and a longitude range of 0° E–45° E, gray lines include all longitudes. Calculations are offset by 150, 90, 30, −30, −90, and −150% for O_3, and by 280, 180, 80, −20, −120, −220% for H_2O, respectively. Different shadings of the colours represent calculations for the SPURT period (2002–2003), the ACE-FTS period (2004–2007), and all years (2002–2007), but note that these lines are hardly distinguishable. Black and gray asterisks denote the number of independent ACE-FTS measurements (N/b) for the restricted and full longitude ranges. Red triangles denote the number of independent SPURT measurements. Dashed lines indicate relative differences of 5% and 15% for O_3 and H_2O, respectively.

levels, hence can be regarded as a "perfect" validation data set (the data points lie outside of the plotted range). When including more longitudes (moving from the black to the gray curves), the standard errors in the mean (i.e. the geophysical variability) for a given number of measurements increase for both tracers, but less so for O_3 than for H_2O. However, when including more longitudes, more ACE-FTS measurements are available (the gain is denoted by the red arrows), and the relative differences are reduced (the gain is denoted by the blue arrows). Table 1 summarizes and Fig. 8 illustrates the results of this evaluation for all altitude levels, indicating the number of available independent measurements of the different datasets, and the theoretically derived standard errors of the mean. The decrease in the standard error is a measure of the benefit obtained by the trade-off. The calculations of the standard errors in the mean when adding more years do not change the curves for the restricted and full longitude ranges significantly, indicating that interannual variations are a negligible source of error. This justifies comparing SPURT with ACE-FTS data, even though the measurements are from different years.

Since ACE-FTS will continue measuring the atmosphere, we will have the possibility to include more measurements in a future comparison which then will improve the error assessment. Or expressed in a different way, one can determine for how many years the ACE-FTS observations should be continued in order to reduce the uncertainty in the climatology below a certain level. For example, in order to obtain uncertainties of 5% for O_3 and 15% for H_2O, respectively, 4–8 additional years of operation of the ACE-FTS would be required. However, we need also to find larger but still reliable data sets for comparison, since the given number of SPURT measurements ultimately limits the achievable improvement, especially in the troposphere and for H_2O. It follows from this evaluation that any measurement errors derived in the comparison with SPURT have to be regarded as upper limits.

We move on to explain our findings in Figs. 7 and 8 in a more theoretical way. The errors in the measurements derived by the suggested method include the variance introduced by geophysical variability σ_g^2, and the variance of the actual measurement error σ_m^2. For the full longitude range the total squared error (ϵ_f) is then written as

$$\epsilon_f = \frac{\sigma_g^2 + \sigma_m^2}{(N/b)}. \quad (3)$$

Table 1. Number of available independent measurements and theoretically derived relative differences (see text for details) for the ACE-FTS and SPURT data at different levels. All are for 40° N–60° N and DJF 2004–2007.

dz from tp	−5.5 km		−2.5 km		−0.5 km		0.5 km		2.5 km		5.5 km	
	N/b	Δ_{rel}	*N/b*	Δ_{rel}	*N/b*	Δ_{rel}	*N/b*	Δ_{rel}	*N/b*	Δ_{rel}	*N/b*	Δ_{rel}
(a) O_3												
SPURT	6	9%	15	11%	42	6%	109	3%	16	8%	–	–%
ACE (45°)	1	38%	2	29%	2	29%	3	18%	3	16%	4	18%
ACE (360°)	2	24%	12	15%	13	13%	23	7%	34	6%	33	7%
(b) H_2O												
SPURT	2	55%	15	26%	42	15%	94	7%	16	15%	–	–%
ACE (45°)	1	133%	2	66%	2	67%	3	40%	3	33%	5	5%
ACE (360°)	2	74%	12	43%	12	46%	23	28%	34	18%	33	3%

For the restricted longitude range, the squared error (ϵ_r) changes to

$$\epsilon_r = \frac{\alpha\sigma_g^2 + \sigma_m^2}{\beta(N/b)}, \tag{4}$$

where α is the ratio of the geophysical variabilities, and β the ratio of the number of measurements between the restricted and the full fields. Both α and β are less than unity for a restricted range. The issue is whether an expanded range leads to reduced errors. This is a trade-off between the effects of α and β. We now can define a trade-off factor γ

$$\gamma = \frac{\epsilon_r}{\epsilon_f} = \frac{\alpha\sigma_g^2 + \sigma_m^2}{\beta(\sigma_g^2 + \sigma_m^2)}. \tag{5}$$

If $\gamma > 1$ then full sampling is preferable to restricted sampling. With

$$\mu = \frac{\sigma_m^2}{\sigma_g^2}, \tag{6}$$

Eq. (5) yields

$$\gamma = \frac{\alpha + \mu}{\beta(1 + \mu)}. \tag{7}$$

If $\mu \ll 1$ (i.e. geophysical variability dominates over measurement error), then $\gamma > 1$ if $\alpha > \beta$, i.e. if the geophysical variance increases less rapidly than the number of measurements as the range of longitudes increases. But for μ=O(1), the condition is less stringent. This is shown in Fig. 9, where we see that, except for H_2O in the case of small measurement errors, this requirement is true and it is beneficial to include all data. This result is also consistent with what we expect from Fig. 8 and Table 1, where moving from the restricted to full longitude range is more beneficial in the case of O_3 than in the case of H_2O.

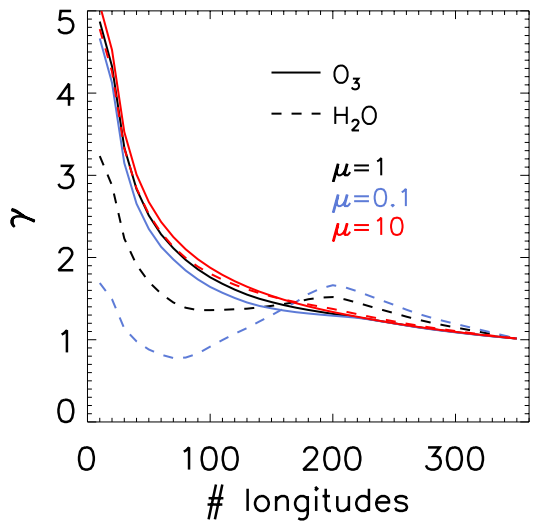

Fig. 9. Trade-off factor γ as a function of number of longitudes included in the climatologies of O_3 (solid lines) and H_2O (dashed lines). For the calculation of the trade-off factor, CMAM data from the altitude bin 1 km below the tropopause and over 5 years of Northern Hemisphere winter data were used. Different colours indicate calculations performed for different μ's, which represent the ratio between the variance of the measurement error and the variance introduced by geophysical variability.

4 Results and discussion

4.1 SPURT versus ACE-FTS tracer-tracer correlations

Figure 10 shows the CO-O_3 and H_2O-O_3 correlations for the ACE-FTS and the SPURT measurements for the latitude range between 30° N and 90° N, full longitude range, and different seasons. ACE-FTS data were accumulated over the years 2004–2007. The SPURT data were obtained during two campaigns in two different years (2002/2003). The agreement between the correlations of the two data sets is remarkable, despite the fact that the measurements were

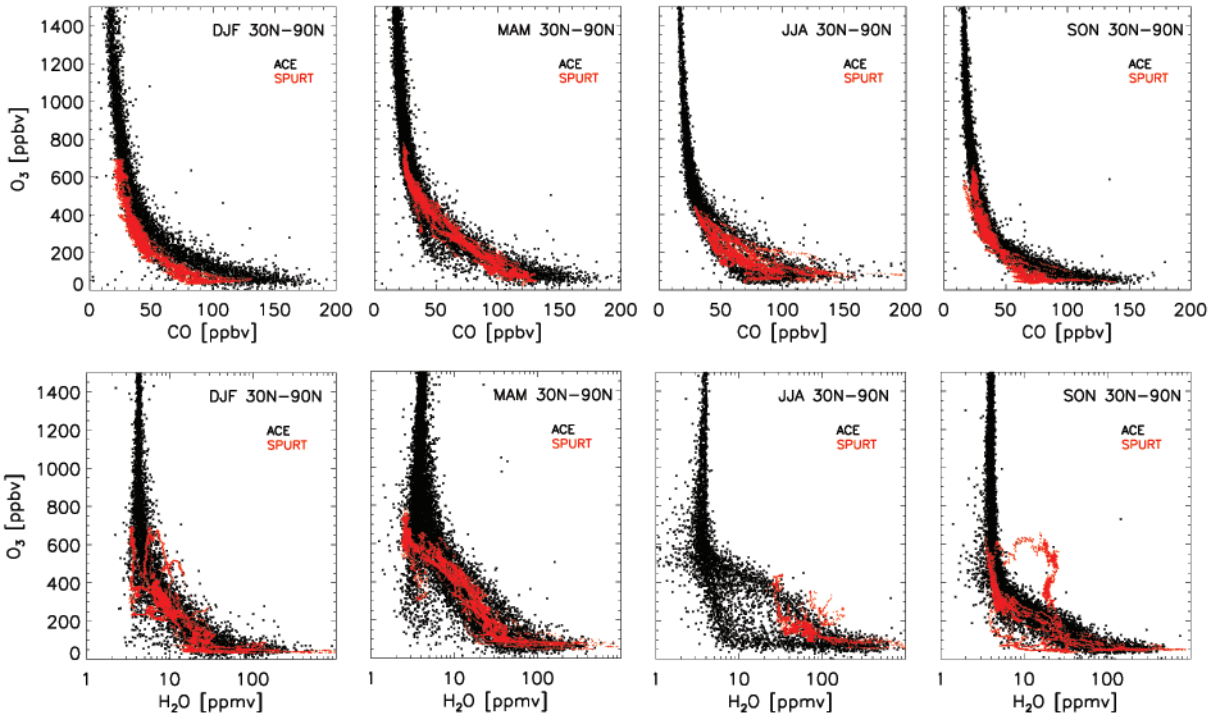

Fig. 10. CO-O_3 (upper panels) and H_2O-O_3 correlations (lower panels) for ACE-FTS satellite (black) and SPURT aircraft (red) data between 30° N and 90° N. From left to right: winter, spring, summer, and autumn measurements. Note that the two data sets were obtained during different years.

carried out in different years. In general, the SPURT data lie well within the range of the ACE-FTS data, and the limited spread (or scatter) of the ACE-FTS indicates a very good precision and vertical resolution of the instrument. Some features such as the "high-heel" shape found in the spring H_2O-O_3 correlation, or the relatively broad transition between the troposphere and the stratosphere in the summer CO-O_3 correlation, are evident in both data sets. With the lower vertical resolution of a satellite instrument, the transition between the stratospheric and the tropospheric branch in the tracer-tracer correlation is expected to be less sharp. Using ACE-FTS data from a narrower latitude range (e.g. between 40° N and 60° N) does not change the results of this evaluation in a significant way, it just tends to decrease slightly the spread in the ACE-FTS correlation. Note that the ceiling altitude of the aircraft was 14 km, hence the largest O_3 values seen in the ACE-FTS were not sampled. On the other hand, the ACE-FTS reaches on average only down to an altitude of 5 km which might explain why the largest H_2O values in the aircraft measurements were never sampled by the satellite. The aircraft measurements in autumn which lie outside the expected H_2O-O_3 correlation range stem from an extraordinary troposphere-to-stratosphere exchange event associated with overshooting deep convection (Hegglin et al., 2004). Such events are mixed into the background within a couple of days and therefore are not a characteristic of the climatology.

4.2 SPURT versus ACE-FTS vertical profiles

The validation of seasonal ACE-FTS CO, O_3, and H_2O data using vertical profiles relative to the tropopause height is presented in Fig. 11a, c and Fig. 12a, c. In order to minimize the uncertainties caused by a potential latitudinal dependence of the mean tracer profiles, we use data between 40° N and 60° N only. Based on the evaluation in Sect. 3.3, we use ACE-FTS data obtained within the full longitude range between 180° W and 180° E. In the case of H_2O, we also provide the evaluation for the restricted longitude range between 45° W and 45° E for comparison. Summer H_2O data are not shown, since the dynamical range of the SPURT measurements did not allow a meaningful comparison (see Fig. 10, third panel of the H_2O-O_3 correlation). The left panels of Figs. 11 and 12 show the mean vertical profiles and their standard deviations for the SPURT (black) and the ACE-FTS data (gray). The right panels show the relative differences and their uncertainties.

The comparison between satellite and aircraft data yields the best result for CO. Average relative differences of the mean are around ±9% in the UT, and ±12% in the LS. This result is comparable to the 16% relative difference derived from a validation using MOZAIC aircraft data in the UTLS by Clerbaux et al. (2007). For O_3, the derived relative differences in the LS are ±19% on average, but closer to the

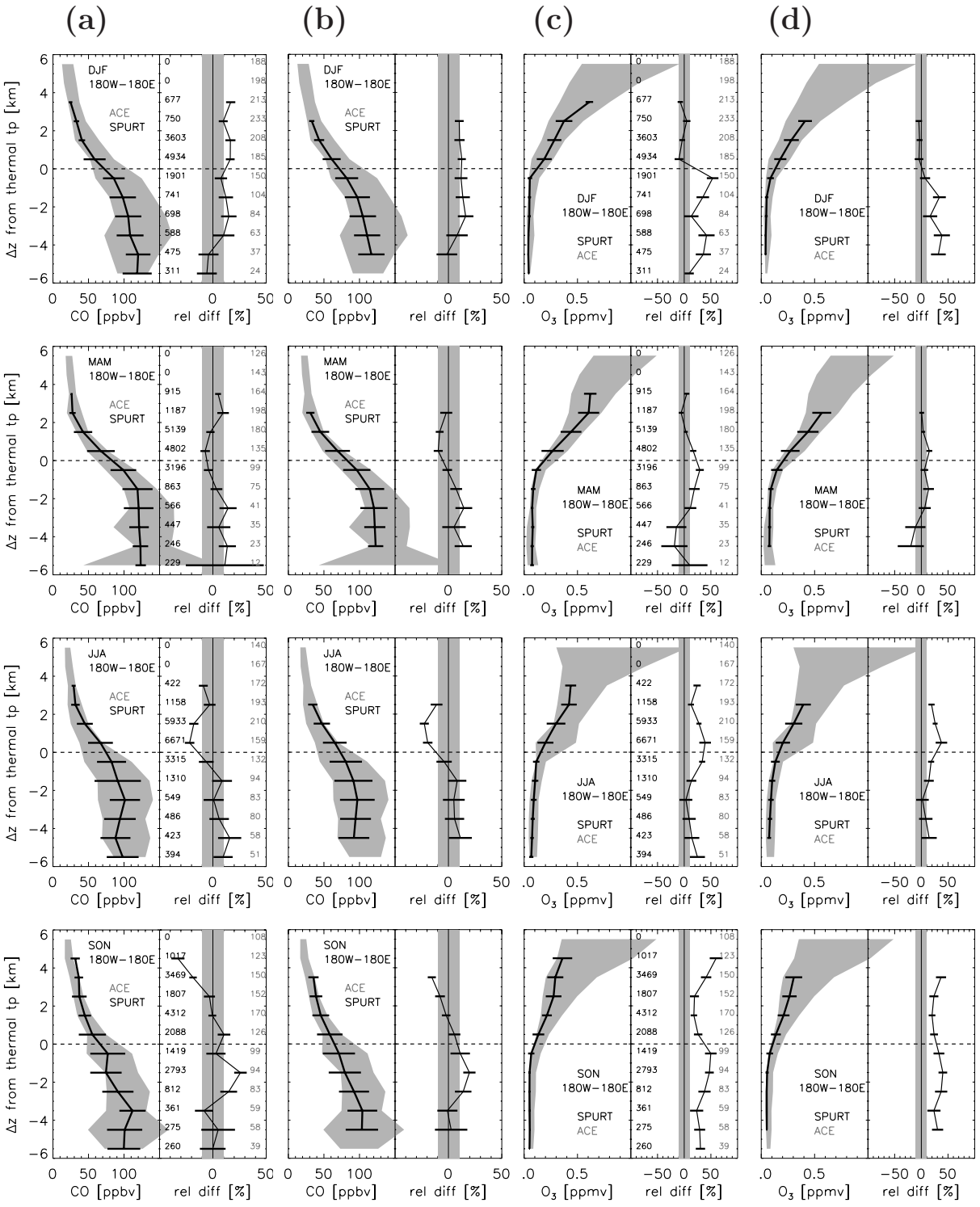

Fig. 11. Comparison of ACE-FTS and SPURT CO and O_3 measurements. Left panels in **(a, c)**: Vertical profiles of mean CO and O_3 mixing ratios with standard deviations (1σ) as a function of altitude relative to the thermal tropopause for SPURT (black) and ACE-FTS (gray), and for DJF, MAM, JJA, and SON (from top to bottom), respectively. Left panels in **(b, d)**: Same as in (a, c) but the SPURT vertical CO and O_3 profiles are smoothed according to Eq. (8) in order to account for the limited vertical resolution of the ACE-FTS instrument. Right panels in (a, b, c, d): Relative differences between ACE-FTS and SPURT mean profiles (black line, calculated using Eq. (1) in text), with positive/negative values indicating a high/low bias in the ACE-FTS measurements. The gray bar indicates ±10% relative difference. Horizontal bars show the uncertainties in the relative differences. The total numbers of measurements are indicated for each altitude bin and data set.

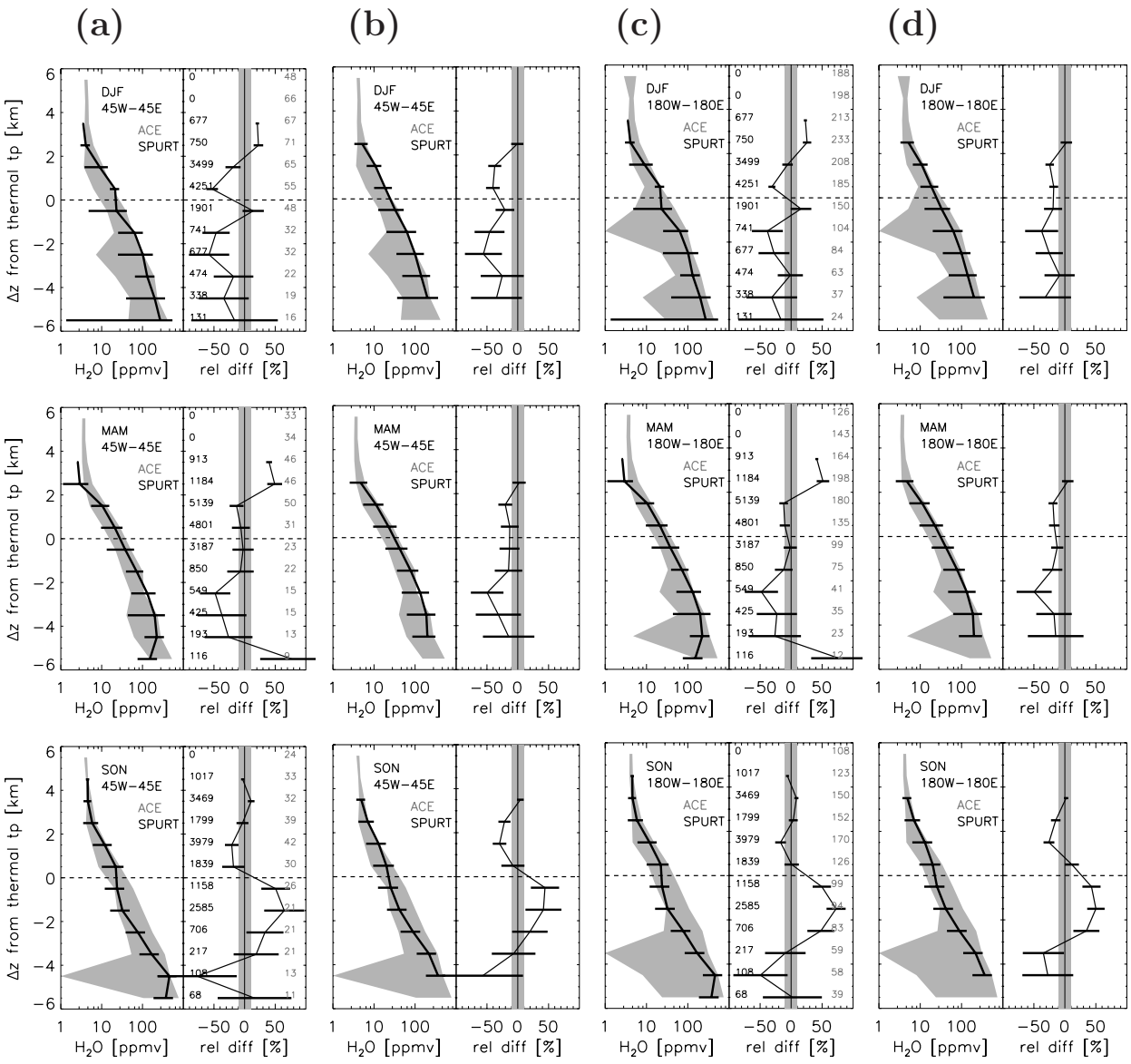

Fig. 12. Same as previous Fig. 10 but for H_2O. **(a, b)** ACE-FTS profiles used to calculate the mean profiles were taken between 45° W and 45° E, and **(c, d)** between 180° W and 180° E. From top to bottom: DJF, MAM, and SON. JJA comparison is omitted due to a too small dynamical range of the SPURT H_2O measurements.

5±5% derived by Dupuy et al. (2008) for altitudes higher than 15 km if the autumn values between 3 and 6 km above the tropopause are excluded from the calculation of the average. This is justified by the evaluation in Fig. 4d, which suggests not including these altitude levels into a comparison. However, in the UT, the relative differences increase to around 25% indicating a persistent high bias of the ACE-FTS O_3. This is consistent with a high bias in the ACE-FTS O_3 interpolated to the tropopause level with respect to MLS, and other solar occultation instruments as shown by Manney et al. (2007). The reason for this bias might be a sensitivity problem of the ACE-FTS when it comes to detecting and accurately measuring low mixing ratios below atmospheric layers with very high mixing ratios. Furthermore, at lower altitudes, fewer microwindows are being used for the O_3 retrieval and there are potentially more interfering species. Indeed, the sensitivity problem is reflected in larger relative retrieval errors in the troposphere. H_2O shows the largest relative differences of about ±18% in the LS, and ±30% in the UT, with the uncertainties of the relative differences suggesting a systematic low bias of the ACE-FTS measurements, at least for the middle to upper troposphere in winter and spring. However, H_2O is the most variable tracer in the troposphere among the species presented here. Indeed the

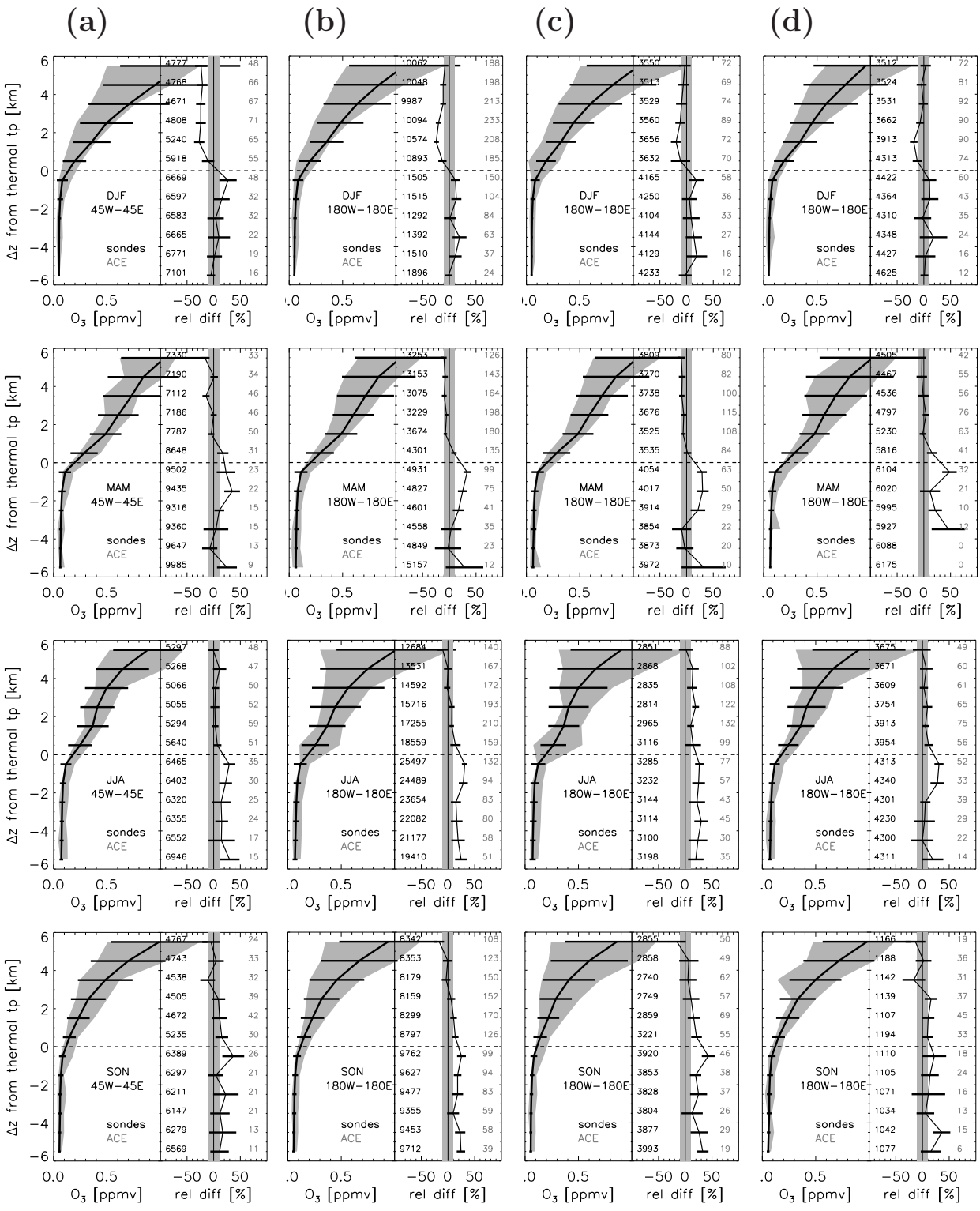

Fig. 13. Comparison of ACE-FTS O_3 and ozonesonde data. From top to bottom: DJF, MAM, JJA, and DJF. Left panels: vertical mean profiles $\pm 1\sigma$ standard deviations for ACE-FTS (in gray) and the ozonesonde data (in black). Right panels: Relative differences between the two data sets and their uncertainties. The total numbers of measurements are indicated for each altitude bin and data set. Profiles were taken **(a)** between 45° W and 45° E and during the years 2004–2007, **(b)** between 180° W and 180° E and during the years 2004–2007, **(c)** same as in (b) but during the year 2005, **(d)** same as in (b) but during the year 2006.

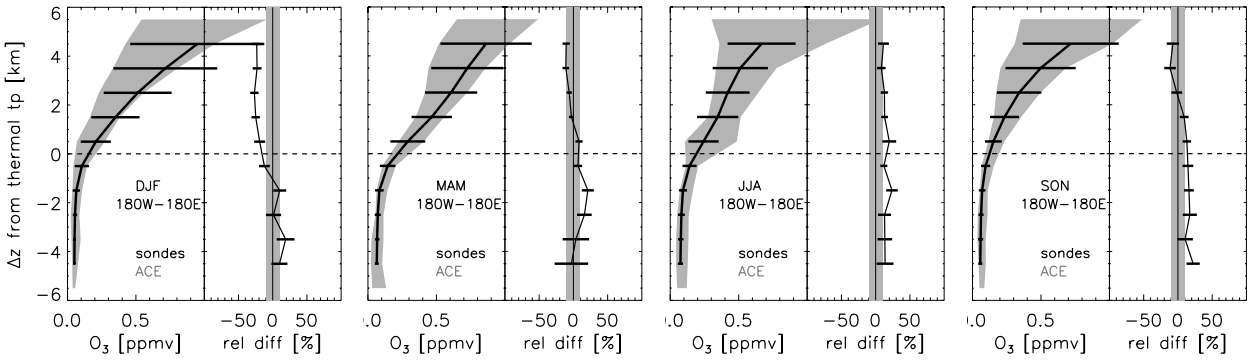

Fig. 14. Same as previous Fig. 13b but with ozonesonde data smoothed according to Eq. (8).

evaluation in Sect. 3.3 indicates that up to 20% of these errors might still be due to undersampling despite the improvement when moving from the restricted to the full longitude range. Including future measurements of the ACE-FTS into this comparison will likely help to decrease the errors further, however, the number of available SPURT measurements ultimately will set another limitation. Hence we will need to find other reliable reference data sets for further comparisons.

While the vertical resolution of the ACE-FTS is affected by its field-of-view, as shown in Fig. 3 there is often a significant altitude oversampling in the vicinity of the tropopause, which may increase the effective vertical resolution of the measurements. This is also suggested by Fig. 10. If the resolution was not better than 3 km, then the relatively sharp transition between the tropospheric and the stratospheric branch of the correlation would be smeared out to a much greater extent. However, at the level of 1 km, there is additionally an inherent smoothing in the retrieval process, since forward model calculations employ data on a 1-km grid. To account (very roughly) for the limitations in the ACE-FTS vertical resolution, we smoothed the SPURT profiles with a triangular function. We apply the following smoothing

$$\chi[i]=0.25\times\chi[i-1]+0.5\times\chi[i]+0.25\times\chi[i+1] \quad (8)$$

where i is the altitude level of a given tracer volume mixing ratio (χ). By smoothing the SPURT O_3, CO, and H_2O profiles with the above equation, we can gauge whether smearing effects from the ACE-FTS vertical resolution have a significant effect on the results.

Figures 11b, d and 12b, d show the relative difference between the smoothed SPURT and the ACE-FTS mean tracer profiles. The comparison with the unsmoothed vertical mean profiles suggests that the smoothing effect accounts only for about 5–15% of the total measurement error, with largest impact just above and below the tropopause. Indeed, largest impact is expected to be found close to the tropopause where strong gradients in the tracer mixing ratios are present. Accordingly, CO seems to be the tracer which is least impacted by the smoothing effect since it exhibits the weakest gradient in tracer mixing ratios across the tropopause. This evaluation suggests that the ACE-FTS is capable of resolving the transition between the troposphere and the stratosphere with high accuracy, and points towards a vertical resolution of around 1 km.

4.3 Ozonesonde versus ACE-FTS vertical profiles

In order to gain more confidence in the validation results obtained in the comparison with aircraft data, which may suffer from the limited number of independent measurements, we also use ozonesonde data which according to our evaluation in Sect. 3.3 represent a "perfect" data set for a non-coincident comparison. We look at different time and spatial samples in order to explore, in an empirical way, how sampling issues might influence the validation results, and if these results are consistent with what we expect from theory (Sect. 3.3).

In Fig. 13a, we show the seasonal comparison between the ozonesonde and the ACE-FTS O_3 data obtained during the full observation period 2004–2007 and restricted to a longitude band between 45° W and 45° E. In general, the results obtained are comparable to the results found from the validation using the SPURT data set. The data show good agreement in the LS, and an ACE-FTS high bias in the UT. However, the noise between the altitude levels is largely reduced, related to the increase in number of measurements when using ozonesonde data instead of SPURT data. In Fig. 13b, the comparison is shown between the data of the two data sets taken within the full longitude range. This increases the number of measurements to compare with by a factor of 2 or 3 for the ozonesondes, and by a factor of 4 for ACE-FTS. Including more measurements leads to a reduction in the relative differences but also in their uncertainties, hence this comparison yields a statistically more meaningful error assessment than using the restricted longitude range. The relative differences between the ACE-FTS O_3 and ozonesonde measurements are ±18% and ±8% for the UT and LS, respectively. The ±8% in the LS is consistent with the results of Dupuy et al. (2008). Finally, we compare the data obtained during

the years 2005 and 2006 separately (Fig. 13c and d, respectively). The derived relative differences of the measurements remain basically unchanged. A change can however be seen in the uncertainties in the relative differences, which increase due to the decrease in the available number of measurements. This is consistent with our earlier analysis based on CMAM and shown in Fig. 8, suggesting that inclusion of ozone measurements taken during different years does not add a discernible amount of geophysical variability.

The validation using SPURT aircraft and ozonesonde data thus yields consistent results, and the uncertainties are seen to follow the theoretical expectations derived in Sect. 3.3. In particular, the measurement errors are better characterized when including measurements obtained over larger spatial and temporal scales, despite the increase in geophysical variability. From this it follows that the trade-off factor (γ) is also >1 in the real atmosphere.

In Fig. 14, we finally show for each season the comparison between the ACE-FTS O_3 data and the ozonesonde profiles smoothed according to Eq. (8) using all available years and the full range of longitudes. The smoothing effect leads to a decrease in the relative differences which is again most pronounced just below the tropopause, and seen to be 5% larger than in the comparison with the SPURT data set. The larger decrease might reflect the additional bias introduced by calculating tropopauses for the ozonesondes from simultaneously measured temperature profiles instead of from the GEOS-4 analyses.

5 Conclusions

In this study we present the validation of ACE-FTS CO, O_3, and H_2O measurements from Canada's SCISAT-1 satellite in the UTLS using SPURT aircraft and ozonesonde measurements. This contributes to the validation efforts of the ACE-FTS data published in this special issue. Other studies (except the one focusing on CO), however, mainly focus on validation in the stratosphere and the mesosphere. In the UTLS, validation of chemical trace gas measurements is a challenging task due to small-scale variability in the tracer fields, strong gradients of the tracers across the tropopause, and scarcity of measurements suitable for validation purposes. We here suggest two alternative/complementary methods for the validation of satellite measurements in the UTLS which are increasingly being used for the validation of chemistry transport models or chemistry climate models: tracer-tracer correlations, and vertical profiles relative to tropopause height. These methods are known to reduce geophysical variability, and thereby provide an "instantaneous climatology", which avoids the need for coincident measurements. The climatological comparison allows one to include all available measurements, not just coincident ones, and thus obtain better statistics and more reliable information about the instrument accuracy and precision.

We first assessed the performance and limitations of the new validation methods. The most important question is whether spatial and temporal variations of the "climatology" are smaller than the errors one is trying to identify. This is quantified through a trade-off factor. Using CMAM data to simulate real measurements, we show that the uncertainty in climatologies due to limited sampling is small enough to allow for a meaningful error assessment in the UTLS if enough measurements are available to describe the climatology. However, this is not always the case, especially for lower altitude levels. For the ACE-FTS, we show that including more measurements obtained over a wider longitude range improves the error assessment, despite the increased geophysical variability. Given the number of currently available measurements, the uncertainty of the ACE-FTS and SPURT climatologies is estimated to be around 5–10% for O_3, and 10–25% for H_2O depending on the altitude level. In contrast, standard validation techniques indicate that ACE-FTS measurement errors in the UTLS region (i.e. below 100 hPa or 15 km) may be as large as 50% or not determinable (Dupuy et al., 2008; Carleer et al., 2008). Hence the error assessment of the ACE-FTS can be improved significantly compared to standard validation techniques using the new methods.

The theoretical analysis also allows one to estimate the number of measurements needed to reduce the uncertainties in the climatologies below a certain level. For example, to obtain uncertainties of $\pm$5% for O_3 and $\pm$15% for H_2O, around 50–100 and 80–120 independent measurements would be needed, respectively, depending on the altitude level. These numbers would require an additional 4–8 years of operation of the ACE-FTS instrument.

From the ozonesonde measurements we determined that using tropopause coordinates is only beneficial close to the tropopause, i.e. around 5(4) km above(below) the tropopause in winter and spring, and around 3(2) km in summer and autumn, respectively. On the other hand, the correlation method can be used wherever correlations are compact, i.e. for tracers whose lifetimes are longer than the transport timescales involved. This is true for both O_3-CO, and O_3-H_2O in the UTLS. In the stratosphere, the method would be applicable to correlations involving N_2O, CH_4, and NO_y. Although not shown in this study, it will also be possible to extract quantitative information about the accuracy of one tracer, as long as the accuracy of the other tracer is known.

We then applied the new methods to the measurements. We used the tracer-tracer correlations to obtain qualitative information on the precision, and the vertical profiles relative to the tropopause to determine quantitatively the uncertainty of the ACE-FTS measurements.

Using the correlation method, we find that the ACE-FTS measurements reproduce characteristic features found in the SPURT aircraft measurements, and hence offer unprecedented precision for a satellite instrument. The evaluation also suggests a vertical resolution of around 1 km. This

implies that the resolution is not solely determined by the field-of-view, which is around 3 km for the ACE-FTS. It seems more important that the ACE-FTS oversamples the atmosphere, i.e. that the vertical spacing between two retrieval levels is found to be <1 km around tropopause altitudes.

Comparison of vertical profiles in tropopause coordinates yields upper bounds of the relative differences in the mean between the ACE-FTS and SPURT aircraft measurements in the UT and LS. These are found to be ±9% and ±12% for CO, ±30% and ±18% for H_2O, and ±25% and ±19% for O_3, respectively. In the troposphere, H_2O shows a persistent low bias in winter and spring, and O_3 a persistent high bias throughout the year. Comparison with ozonesonde data further improves the error assessment, as predicted by the theoretical derivation of the trade-off factor. Relative differences now are found to be ±18% and ±8% for the UT and LS, respectively, and also their uncertainties decrease relative to the ones obtained in the SPURT evaluation.

When taking into account the smearing effect of the limited spacing between the measurements of the ACE-FTS instrument, the errors decrease by 5–15% around the tropopause. This limitation in the vertical resolution must be taken into account when interpreting the ACE-FTS UTLS measurements. Note that the tropopause heights of the different data sets used in validation studies should preferably be derived using the same algorithms and analyses (e.g. GEOS-4, GEOS-5, MetO (Met Office), or ECMWF (European Centre for Medium-Range Weather Forecasts)), since differences in their characteristics may present an additional error source. The significance of the results presented here, especially for the H_2O evaluations, should be tested with further aircraft, balloon, and satellite data. Nevertheless, the analysis indicates that the ACE-FTS offers unprecedented precision and vertical resolution in the UTLS for a satellite instrument, allowing a new global perspective on UTLS trace gas distributions. Moreover, our study shows that aircraft observations provide valuable data sets for satellite validation, especially for tracers which are not routinely being measured in the atmosphere (e.g. CO). By applying the methods presented here, more value can be extracted from historical data sets. It furthermore allows the comparison of satellites whose observation periods do not overlap, an issue frequently pointed out in the literature.

Acknowledgements. The Atmospheric Chemistry Experiment (ACE), also known as SCISAT, is a Canadian-led mission mainly supported by the Canadian Space Agency and the Natural Sciences and Engineering Research Council of Canada. Work at the Jet Propulsion Laboratory, California Institute of Technology was done under contract with the National Aeronautics and Space Administration. M. Hegglin has been supported by the Canadian Foundation for Climate and Atmospheric Sciences and the Canadian Space Agency through the C-SPARC program. Thanks to the WOUDC for providing ozonesonde data. Special thanks to E. Dupuy for assistance with the ozonesonde data, and to C. McLandress for helpful comments on the manuscript.

Edited by: T. Wagner

References

Beagley, S. R., de Grandpré, J., Koshyk, J. N., McFarlane, N. A., and Shepherd, T. G., Radiative-dynamical climatology of the first-generation Canadian Middle Atmosphere Model, Atmos.-Ocean, 35, 293–331, 1997.

Bernath, F., McElroy, C. T., Abrams, M. C., et al.: Atmospheric Chemistry Experiment (ACE): Mission overview, Geophys. Res. Lett., 32, L15S01, doi:10.1029/2005GL022386, 2005.

Bloom, S. C., da Silva, A. M., Dee, D. P., et al.: The Goddard Earth Observing Data Assimilation System, GEOS DAS Version 4.0.3: Documentation and Validation, NASA, 104606 V26, 2005.

Boone, C. D., Nassar, R., Walker, K. A., et al.: Retrievals for the Atmospheric Chemistry Experiment Fourier Transform Spectrometer, Appl. Opt., 44, 7218–7231, 2005.

Carleer, M. R., Boone, C. D., Walker, K. A., et al.: Validation of water vapour profiles from the Atmospheric Chemistry Experiment (ACE), Atmos. Chem. Phys. Discuss., 8, 4499–4559, 2008, http://www.atmos-chem-phys-discuss.net/8/4499/2008/.

Clerbaux, C., Coheur, P.-F., Hurtmans, D., Barret, B., Carleer, M., Colin, R., Semeniuk, K., McConnell, J. C., Boone, C., and Bernath, P.: Carbon monoxide distribution from the ACE-FTS solar occultation measurements, Geophys. Res. Lett., 32, L16S01, doi:10.1029/2005GL022394, 2005.

Clerbaux, C., George, M., Turquety, S., et al.: CO measurements from the ACE-FTS mission: validation using ground-based, airborne and satellite observations, Atmos. Chem. Phys. Discuss., 7, 15 277–15 340, 2007.

de Grandpré, J., Beagley, S. R., Fomichev, V. I., Griffioen, E., McConnell, J. C., Medvedev, A. S., and Shepherd, T. G., Ozone climatology using interactive chemistry: Results from the Canadian Middle Atmosphere Model, J. Geophys. Res., 105, 26 475–26 491, 2000.

Douglass, A. R., Schoeberl, M. R., Rood, R. B., and Pawson, S.: Evaluation of transport in the lower tropical stratosphere in a global chemistry and transport model, J. Geophys. Res., 108(D9), 4259, doi:10.1029/2002JD002696, 2003.

Dupuy, E., Walker, K. A., Kar, J., et al.: Validation of ozone measurements from the Atmospheric Chemistry Experiment (ACE), Atmos. Chem. Phys. Discuss., 8, 2513–2656, 2008, http://www.atmos-chem-phys-discuss.net/8/2513/2008/.

Engel, A., Bönisch, H., Brunner, D., et al.: Highly resolved observations of trace gases in the lowermost stratosphere and upper troposphere from the Spurt project: an overview, Atmos. Chem. Phys., 6, 283–301, 2006, http://www.atmos-chem-phys.net/6/283/2006/.

Eyring, V., Butchart, N., Waugh, D. W., et al.: Assessment of temperature, trace species and ozone in chemistry-climate model simulations of the recent past, J. Geophys. Res., 111, D22308, doi:10.1029/2006JD007327, 2006.

Fischer, H., Wienhold, F. G., Hoor, P., Bujok, O., Schiller, C., Siegmund, P., Ambaum, M., Scheeren, H. A., and Lelieveld, J.: Tracer correlations in the northern high latitude lowermost stratosphere: Influence of cross-tropopause mass exchange, Geophys. Res. Lett., 27(1), 97–100, doi:10.1029/1999GL010879, 2000.

Froidevaux, L., Livesey, N. J., Read, W. G., et al.: Early validation analyses of atmospheric profiles from EOS MLS on the Aura

satellite, IEEE Transaction on Geoscience and Remote Sensing, 44(5), 1106–1121, 2006.

Froidevaux, L., Jiang, Y., Lambert, A., et al. : Validation of Aura Microwave Limb Sounder stratospheric ozone measurements, J. Geophys. Res., doi:10.1029/2007JD008771, in press, 2008.

Fussen, D., Vanhellemont, F., Dodion, J., Bingen, C., Walker, K. A., Boone, C. D., McLeod, S. D. and Bernath, P. F.: Initial intercomparison of ozone and nitrogen dioxide number density profiles retrieved by the ACE-FTS and GOMOS occultation experiments, Geophys. Res. Lett., 32, L16S02, doi:10.1029/2005GL022468, 2005.

Hegglin, M. I., Brunner, D., Wernli, H., et al.: Tracing troposphere-to-stratosphere transport above a mid-latitude deep convective system, Atmos. Chem. Phys., 4, 741–756, 2004, http://www.atmos-chem-phys.net/4/741/2004/.

Hegglin, M. I., Brunner, D., Peter, T., et al.: Measurements of NO, NO_y, N_2O, and O_3 during SPURT: implications for transport and chemistry in the lowermost stratosphere, Atmos. Chem. Phys., 6, 1331–1350, 2006, http://www.atmos-chem-phys.net/6/1331/2006/.

Hegglin, M. I., and Shepherd, T. G.: O_3-N_2O correlations from the Atmospheric Chemistry Experiment: Revisiting a diagnostic of transport and chemistry in the stratosphere, J. Geophys. Res., 112, D19301, doi:10.1029/2006JD008281, 2007.

Hoor, P., Gurk, C., Brunner, D., Hegglin, M. I., Wernli, H., and Fischer, H.: Seasonality and extent of extratropcial TST derived from in-situ CO measurements during SPURT, Atmos. Chem. Phys., 4, 1427–1442, 2004, http://www.atmos-chem-phys.net/4/1427/2004/.

Jin, J. J., Semeniuk, K., Jonsson, A. I., et al.: Co-located ACE-FTS and Odin/SMR stratospheric-mesospheric CO 2004 measurements and comparison with a GCM, Geophys. Res. Lett., 32, L15S03, doi:10.1029/2005GL022433, 2005.

Krebsbach, M., Brunner, D., Günther, G., Hegglin, M., Maser, R., Mottaghy, D., Riese, M., Spelten, N., Wernli, H., and Schiller, C.: Seasonal cycles and variability of O_3 and H_2O in the UT/LMS during SPURT, Atmos. Chem. Phys., 6, 109–125, 2006, http://www.atmos-chem-phys.net/6/109/2006/.

Manney, G. L., Daffer, W. H., Zawodny, J. M., et al.: Solar occultation satellite data and derived meteorological products: Sampling issues and comparisons with Aura Microwave Limb Sounder, J. Geophys. Res., 112, D24S50, doi:10.1029/2007JD008709, 2007.

Monahan, K. P., Pan, L. L., McDonald, A. J., Bodeker, G. E., Wei, J., George, S. E., Barnet, C. D., and Maddy, E.: Validation of AIRS v4 ozone in the UTLS using ozonesondes from Lauder, NZ and Bolder, USA, J. Geophys. Res., 112, D17304, doi:10.1029/2006JD008181, 2007.

McHugh, M., Magill, B., Walker, K. A., Boone, C. D., Bernath, P. F., and Russell III, J. M.: Comparison of atmospheric retrievals from ACE and HALOE, Geophys. Res. Lett., 32, L15S10, doi:10.1029/2005GL022403, 2005.

Pan, L. L., Randel, W. J., Gary, B. L., Mahoney, M. J., and Hintsa, E. J.: Definitions and sharpness of the extratropical tropopause: A trace gas perspective, J. Geophys. Res., 109, D23103, doi:10.1029/2004JD004982, 2004.

Pan, L. L., Wei, J. C., Kinnison, D. E., Garcia, R. R., Wuebbles, D. J., and Brasseur, G. P.: A set of diagnostics for evaluating chemistry-climate models in the extratropical tropopause region, J. Geophys. Res., 112, D09316, doi:10.1029/2006JD007792, 2007.

Plumb, R. A. and Ko, M. K. W.: Interrelationships between mixing ratios of long-lived stratospheric constituents, J. Geophys. Res., 97, 10 145–10 156, 1992.

Reichler, T., Dameris, M., and Sausen, R.: Determining the tropopause height from gridded data, Geophys. Res. Lett., 30(20), 2042, doi:10.1029/2003GL018240, 2003.

Sankey, D. and Shepherd, T. G.: Correlations of long-lived chemical species in a middle atmosphere general circulation model, J. Geophys. Res., 108(D16), 4494, doi:10.1029/2002JD002799, 2003.

Strahan, S. E., Duncan, B. N., and Hoor, P.: Observationally derived transport diagnostics for the lowermost stratosphere and their application to the GMI chemistry and transport model, Atmos. Chem. Phys., 7, 2435–2445, 2007, http://www.atmos-chem-phys.net/7/2435/2007/.

Stratospheric Processes and their Role in Climate (SPARC): Assessment of trends in the vertical distribution of ozone, SPARC report No.1, WMO Global Ozone Res. and Monit. Project Rep. 43, Editors N. Harris, R. Hudson and C. Phillips, World Meteorol. Org., Geneva, 1998.

WMO: Meteorology – A three-dimensional science, WMO Bulletin, 6, 134–138, 1957.

Walker, K. A., Randall, C. E., Trepte, C. R., Boone, C. D., and Bernath, P. F.: Initial validation comparisons for the Atmospheric Chemistry Experiment (ACE-FTS), Geophys. Res. Lett., 32, L16S04, doi:10.1029/2005GL022388, 2005.