
S. Mieruch¹,*, M. Weber¹, C. von Savigny¹, A. Rozanov¹, H. Bovensmann¹, J. P. Burrows¹, P. F. Bernath², C. D. Boone³, L. Froidevaux⁴, L. L. Gordley⁵, M. G. Mlynczak⁵, J. M. Russell III⁷, L. W. Thomason⁶, K. A. Walker⁷, and J. M. Zawodny⁶

¹Institut für Umweltphysik, Universität Bremen FB1, Bremen, Germany
²Department of Chemistry, University of York, Heslington, York, UK
³Department of Chemistry, University of Waterloo, Waterloo, Ontario, Canada
⁴Jet Propulsion Laboratory, California Institute of Technology, Pasadena, California, USA
⁵NASA Langley Research Center, Hampton, Virginia, USA

*Corresponding author.

© Author(s) 2011. CC Attribution 3.0 License.
Center for Atmospheric Sciences, Hampton University, Hampton, Virginia, USA

Department of Physics, University of Toronto, Toronto, Ontario, Canada

*now at: CM-SAF, Dept. Climate and Environment, Deutscher Wetterdienst (DWD), Offenbach, Germany

Received: 25 July 2011 – Accepted: 28 July 2011 – Published: 2 August 2011

Correspondence to: M. Weber (weber@uni-bremen.de)

Published by Copernicus Publications on behalf of the European Geosciences Union.
Abstract

SCIAMACHY limb scatter ozone profiles from 2002 to 2008 have been compared with MLS (2005–2008), SABER (2002–2008), SAGE II (2002–2005), HALOE (2002–2005) and ACE-FTS (2004–2008) measurements. The comparison is performed for global zonal averages and heights from 10 to 50 km in one km steps. The validation was performed by comparing monthly mean zonal means and by comparing averages over collocated profiles within a zonal band and month. Both approaches yield similar results. For most of the stratosphere SCIAMACHY agrees to within 10 % or better with other correlative data. A systematic bias of SCIAMACHY ozone of up to 100 % between 10 and 20 km in the tropics points to some remaining issues with regard to convective cloud interference. Statistical hypothesis testing reveals at which altitudes and in which region differences between SCIAMACHY and other satellite data are statistically significant. We also estimated linear trends from monthly mean data for different periods where SCIAMACHY has common observations with other satellite data using a classical trend model with QBO and seasonal terms in order to draw conclusions on potential instrumental drifts as a function of latitude and altitude. SCIAMACHY exhibits a statistically significant negative trend in the range of of about 1–3 % per year depending on latitude during the period 2002–2005 (overlapping with HALOE and SAGE II) and somewhat less during 2002–2008 (overlapping with SABER) in the altitude range of 30–40 km, while in the period 2004–2008 (overlapping with MLS and ACE-FTS) no significant trends are observed. The statistically significant negative trends only observed with SCIAMACHY data point at some residual effects from errors in the tangent height registration.

1 Introduction

Ozone profile observations from spaceborne measurements are currently carried out by several instruments on different satellite platforms. Ozone profiles from different instruments have been compared extensively against each other (Bracher et al., 2005; Boyd et al., 2007; Kuttippurath et al., 2007; Dupuy et al., 2009, and citations therein).
SCIAMACHY (Scanning Imaging Spectrometer for Atmospheric Chartography) provides limb scatter measurements of ozone profiles with global coverage since 2002 and this data set has become a valuable data set for assessing long-term trends in combination with other satellite data (Steinbrecht et al., 2009; Jones et al., 2009). The comparison of ozone profiles with correlative data sets is an important subject and contributes to a better understanding of potential errors and systematic biases in the individual data sets. Moreover it contributes to improvements in retrieval algorithms and instrumental designs. Most comparisons have been performed either using selected collocations or monthly mean zonal averages and often only short time windows have been analyzed, which is usually the case when validating shortly after launch. Statistical hypothesis testing is still not used extensively. Thus the evaluation of the validation results by comparison depends on the methods applied, used averages, and time coverage of the correlative data sets.

Many satellite time series of ozone profiles have reached a sufficiently long record, which allows comparisons of up to six years between SCIAMACHY and several other satellite instruments. In this study we will investigate how validation based upon collocated pairs of profiles compares to results from comparing monthly mean zonal mean data directly. The latter method avoids the search for collocated pairs of data that can be time consuming. In this study comparisons are performed for both cases using a latitude/height grid from $-90^\circ$ to $+90^\circ$ in $10^\circ$ steps and heights from $10$ km to $50$ km, except for the SABER comparisons, where the comparisons are limited to altitudes above $25$ km. To judge upon the observed differences and reveal systematic disagreement, we extensively make use of statistical hypothesis testing. The next obvious step in analyzing long time series is the estimation of trends and trend differences between instruments to draw conclusions on potential drifts with time. The trend analysis as commonly done for long-term decadal trend assessments is here solely used for diagnostic purposes to evaluate the quality of the SCIAMACHY data as a function of time with respect to the correlative data.
In this study limb scatter ozone profiles from SCIAMACHY (SCanning Imaging Absorption spectro-Meter for Atmospheric CHartographY) on Envisat, are compared with measurements from MLS (Microwave Limb Sounder) on Aura (since 2004), SABER (Sounding of the Atmosphere using Broadband Emission Radiometry) on TIMED (since 2002), SAGE II (Stratospheric Aerosol and Gas Experiment II) on ERBS (1984–2005), HALOE (HALogen Occultation Experiment) on UARS (1991–2005), and ACE-FTS (Atmospheric Chemistry Experiment-Fourier Transform Spectrometer) aboard SCISAT-1 (since 2004). 

In Sect. 2 we shortly introduce the various satellite instruments and measurement techniques. Section 3 investigates the preprocessing of the different data sets, followed by SCIAMACHY comparisons with other measurements, and finally the statistical hypothesis testing to judge upon the significance of differences. The trend estimation is summarized in Sect. 4, which includes significance tests of the trends and the comparison of trends between independent space borne observations. We summarize and conclude our results in Sect. 5 and give a short outlook.

2 Satellite data sets

2.1 SCIAMACHY on Envisat

SCIAMACHY (SCanning Imaging Absorption spectroMeter for Atmospheric CHartographY) is a passive imaging grating spectrometer on the Environmental Satellite Envisat (launched in March 2002). Envisat flew until end of 2010 in a sun-synchronous orbit at a height of about 800 km and now 17 km lower. It has an orbital period of about 100 min with 14.3 orbits per day. Envisat crosses the equator at 10:00 local time. SCIAMACHY measures the reflected, backscattered and transmitted solar radiation upwelling from the top of atmosphere (Burrows et al., 1990, 1995; Bovensmann et al., 1999; Gottwald et al., 2006; Gottwald and Bovensmann, 2011). It covers the spectral region from 214 nm to 2386 nm with some gaps in the near IR in eight spectral
channels. Atmospheric trace gases such as \( \text{O}_3 \), \( \text{NO}_2 \), \( \text{O}_2 \), \( \text{H}_2\text{O} \), \( \text{CH}_4 \), \( \text{CO}_2 \), \( \text{CO} \), \( \text{BrO} \), \( \text{OCIO} \), \( \text{SO}_2 \) and \( \text{IO} \) as well as clouds and aerosols can be retrieved with SCIAMACHY (Bovensmann et al., 2011).

SCIAMACHY has three viewing modes: nadir, solar and lunar occultation, and limb. The limb mode permits the near-global retrieval of vertical profiles of trace gases in the sunlit part of the earth. The vertical field of view in limb observation mode is about 2.6 km at the tangent point. The horizontal resolution of the limb measurements is usually 240 km across track and about 400 km along track. The vertical scanning is done in 3.3 km steps. Due to the alternating limb-nadir sequences in nominal operation and the complete horizontal scan of 960 km across-track, coverage of the complete sunlit earth is achieved within six days (Bovensmann et al., 1999).

In this study we use ozone profile data retrieved in the SCIAMACHY limb view mode. The retrieval version 2.1 as described in Sonkaew et al. (2009) is an extension of the initial retrieval scheme by von Savigny et al. (2005b) and combines three wavelengths (525, 600 and 675 nm) in the \( \text{O}_3 \) Chappuis band and several wavelengths in the UV Hartley-Huggins band in order to cover altitudes from above the tropopause up to about 70 km. The inversion is done using a non-linear, iterative optimal estimation approach together with the SCIATRAN radiative transfer model (RTM) (Rozanov et al., 2005). The vertical resolution of the SCIAMACHY limb scatter ozone profiles is about 4 km.

### 2.2 SAGE II on ERBS

SAGE II (Stratospheric Aerosol and Gas Experiment II) (Mauldin III et al., 1985) on ERBS (Earth Radiation Budget Satellite) was launched in October 1984 and performs two occultation measurements per orbit (sunrise and sunset). This results in about 30 ozone profiles per day. With the SAGE II instrument, ozone, nitrogen dioxide, water vapor, and aerosols can be retrieved using seven channels between 0.385 \( \mu \text{m} \) and 1.02 \( \mu \text{m} \). Ozone is retrieved from the 0.603 \( \mu \text{m} \) channel measuring in the Chappuis ozone band (Chu et al., 1988). The current retrieval version as used for the SCIAMACHY validation is 6.2 (Thomason et al., 2004).
The agreement between SAGE II version 6.1 ozone and ozonesondes was shown to be approximately within 10% down to the tropopause (Wang et al., 2002).

### 2.3 HALOE on UARS

Similar to SAGE II, HALOE (HALogen Occultation Experiment) (Russell III et al., 1993) onboard UARS (Upper Atmosphere Research Satellite) (Reber et al., 1993) is a solar occultation instrument, performing about fifteen sunrise and fifteen sunset measurements per day. It has the second largest data record with nearly 15 years of measurements between 1991 and 2005. The HALOE instrument uses eight channels from 0.245 µm to 10.04 µm measuring among others ozone, nitrogen oxide, nitric oxide, water vapor, and methane. Ozone is retrieved from the 9.6 µm channel with an onion-peeling scheme (Russell III and Drayson, 1972).

Comparisons between HALOE V17 data and other correlative ozone measurements showed agreement to within 5% from roughly 20 km up to 50 km altitude and a stronger low bias in the lowermost stratosphere (Brühl et al., 1996). Direct comparisons between SAGE II and HALOE V19 data showed an agreement to within 5–10% throughout the stratosphere with a tendency of HALOE being slightly lower than SAGE II (Morris et al., 2001; Nazaryan et al., 2005; Terao and Logan, 2007). In this study HALOE V19 data is used.

### 2.4 SABER on TIMED

The SABER (Sounding of the Atmosphere using Broadband Emission Radiometry) instrument was launched in December 2001 onboard the TIMED (Thermosphere Ionosphere Mesosphere Energetics and Dynamics) satellite (Russell III et al., 1999; Remsberg et al., 2008). TIMED is in a non sun-synchronous orbit and the equator crossing time shifts every day by approximately twelve minutes. This enables, for example, studies of diurnal variability in mesospheric ozone (Huang et al., 2008; Dikty et al., 2010a). SABER is an infrared spectrometer, measuring from 1.27 µm to 16.9 µm in limb mode.
Ozone profiles are retrieved from the 1.27 µm airglow (Mlynczak et al., 2007) during daytime and from 9.6 µm rotational vibrational transitions (Rong et al., 2009) during day and night. SABER measures perpendicular to the flight direction of TIMED and the satellite platform performs a yaw maneuver approximately every 60 days, which results in reversing the scanning direction by 180°. One consequence of this behavior is a drastic change in the local time of measurement. This can be seen in the differences between SCIAMACHY and SABER monthly mean ozone profiles derived from the daytime data. Therefore, we restrict our comparisons to the nighttime data from the version 1.07 thermal IR retrieval.

So far, the only SABER ozone validation was reported by Dupuy et al. (2009) showing a general agreement of SABER with ACE-FTS to within 7% between 25 km and 50 km altitude based upon SABER version 1.06 data retrieved from thermal emissions in the 9.6 µm band. Below 25 km the zonal mean monthly means from SABER show significant larger standard deviations than other satellite so that the comparison in this paper will be limited to altitudes above 25 km. Poleward of 55° SABER samples only 6 months out of the year, alternating every 2 months between hemispheres. The comparisons were therefore limited to latitudes smaller than 50°.

2.5 ACE-FTS on SCISAT-1

ACE-FTS (Atmospheric Chemistry Experiment-Fourier Transform Spectrometer) is a Fourier transform spectrometer making measurements at high spectral resolution (0.02 cm⁻¹) in the spectral range between 2.2 µm and 13 µm (Bernath et al., 2005). ACE-FTS was launched on SCISAT-1 in August 2003. It makes solar occultation observations and ozone is retrieved from the thermal infrared absorption spectra (9.6 µm). An extensive validation of the ACE-FTS “V2.2 Update” ozone profiles was conducted by Dupuy et al. (2009). This work showed a 0–10% positive bias with respect to the majority of correlative data sets from other satellites, ground-based, and balloon-borne observations between 16 and 44 km altitude. In the upper atmosphere (45–60 km) the positive bias becomes larger, on average 20% (Dupuy et al., 2009).
2.6 MLS on Aura

MLS (Microwave Limb Sounder) consists of seven radiometers, which cover five broad spectral regions from 118 GHz to 2.5 THz. MLS onboard the Aura satellite (Schoeberl et al., 2006) was launched in July 2004 into a sun-synchronous quasi-polar orbit. It operates at an altitude of \( \approx 700 \) km, the orbit inclination is 98° and the ascending node crossing takes place at 13:45 local time. Detailed information on the instrument, calibration and performance can be found in Jarnot et al. (2006), Cofield and Stek (2006), and Pickett (2006). MLS provides a very good global sampling with about 3500 profiles per day including day- and night-time measurements. Retrievals of seventeen primary atmospheric parameters such as pressure, temperature, and fourteen trace gas constituents including ozone and water vapor are possible (Waters et al., 2006). The ozone retrieval is based on the Optimal Estimation Method (Rodgers, 2000) and further details can be found in Livesey et al. (2006).

For most part of the stratosphere MLS has a precision of about 5% and 10% in the lowermost stratosphere (Froidevaux et al., 2008). From comparisons with other satellite data, ozone sondes, groundbased lidar and microwave radiometers mean average differences of MLS v.2.2 ozone data with respect to the correlative data are generally within 8% above 150 hPa (Hocke et al., 2007; Jiang et al., 2007; Froidevaux et al., 2008).

3 Ozone profile comparisons

The long term comparison of SCIAMACHY limb ozone profiles with SAGE II, HALOE, SABER, ACE-FTS, and MLS follows two approaches. On the one hand we compare monthly mean zonal mean data and on the other hand we analyze monthly mean differences of collocated data within a zonal band. In both cases the comparison is performed on a latitude-height grid from \(-90°\) to \(+90°\) in 10° steps and from 10 km to 50 km altitude in one km step, except for SABER, where we limited the range to
25–50 km. Special emphasis is placed on statistical significance of differences between instruments by applying hypothesis tests.

3.1 Data preprocessing

The SCIAMACHY limb ozone data are retrieved as number densities, i.e. molec cm$^{-3}$, and are available on a 10 km to 70 km height grid in 1 km intervals. All other correlative data have been interpolated to the same altitude grid and converted to number density if only available as volume mixing ratios (vmr).

MLS data are available as vmr on a pressure grid, thus we had to transform the pressure data first to the height grid using first the retrieved geopotential height (GPH) data from MLS (Schwartz et al., 2008) and then convert GPH to geometrical altitude using standard approximations. The MLS temperature data (Schwartz et al., 2008) were also used to convert ozone vmr into number densities. Finally we linearly interpolated the ozone data to the SCIAMACHY altitude grid.

SABER data are available as mixing ratios and have been converted to concentrations using the provided temperature and pressure data in the SABER data product and were then interpolated to the SCIAMACHY altitude grid.

SAGE II data are available as concentrations on a 0.5 km grid, thus they had only to be regridded to the SCIAMACHY 1 km grid. HALOE ozone profiles are available on a 0.3 km height grid as mixing ratios. They have been converted to number density units using temperature and pressure data provided in the data products and interpolated to the SCIAMACHY altitude grid. ACE-FTS ozone data were available on a 1 km grid as volume mixing ratios and the ACE-FTS retrieved temperatures were used for conversion.

Two profiles are considered collocated when the distance is less than 400 km and time difference between the collocated measurements is less than 4 h. For SABER measurements where we used only nighttime data we relaxed the maximum time difference to 24 h. In addition the time criterion for MLS was relaxed to eight hours to achieve a sufficient number of collocations at southern mid-latitudes. For each instrument
pair with SCIAMACHY all collocations during the common observation period were searched for and used in this study.

### 3.2 Intercomparison

For the comparison of monthly mean zonal mean data we first generated the binned latitude/height/month \((\ell, h, m)\) data sets for each instrument. That means, we averaged over all ozone profiles \(u_i\) within the \((\ell, h, m)\) bin:

\[
U_{\ell, h, m} = \frac{1}{K} \sum_{i=1}^{K} u_i.
\]

As an example monthly mean zonal mean ozone profiles from SCIAMACHY, MLS, SABER, SAGE II, HALOE, and ACE-FTS in the zonal band 20° S–30° S and from August 2004 are shown in Fig. 1. The relative differences in percent between SCIAMACHY monthly means \((U)\) and the other instruments \((V)\) for each \((\ell, h, m)\) bin are calculated as follows

\[
D(U, V)_{\ell, h, m} = \frac{2 \cdot (U_{\ell, h, m} - V_{\ell, h, m})}{(U_{\ell, h, m} + V_{\ell, h, m})} \cdot 100.
\]

The right panel of Fig. 1 shows the differences between SCIAMACHY zonal mean profiles and the other instruments for August 2004 in the 20° S–30° S zonal band. This month is one of the few months where data from all satellite data used in this study were available. The agreement above 20 km altitude in this case is generally within 10 %.

The mean differences \(D_{\ell, h, m}\) between two instruments over all commonly observed \(N\) months is given by

\[
\frac{D(U, V)_{\ell, h}}{N} = \frac{1}{N} \sum_{m=1}^{N} D(U, V)_{\ell, h, m}.
\]
and are shown in Fig. 2 for all satellites with respect to SCIAMACHY. In addition the bottom right panel of Fig. 2 shows the mean difference profiles averaged over the latitude range 60° S–60° N.

We also made a comparison limited to zonal mean of collocated profile data, normally a subset $i$ of all profiles used in the zonal mean calculation. The percent differences between each individual pair $(u, v)$ of collocated observations were computed, i.e.

$$D(u, v)_{\ell,h,m,i} = \frac{2 \cdot (u_{\ell,h,m,i} - v_{\ell,h,m,i})}{(u_{\ell,h,m,i} + v_{\ell,h,m,i})} \cdot 100,$$

and the mean relative differences over all $N$ months of common observations for each $(\ell, h)$ bin were derived as follows

$$D(u, v)_{\ell,h} = \frac{1}{N} \sum_{m=1}^{N} \frac{1}{K} \sum_{i=1}^{K} D(u, v)_{\ell,h,m,i},$$

where $K$ denotes here the number of collocated data found in an $(\ell, h, m)$ bin. The results are presented in Fig. 3, again the bottom right panel depicts the global mean difference profiles from 60° S–60° N.

Table 1 summarizes the results. Shown are the global differences (60° S–60° N) of SCIAMACHY to the other satellite data for several altitude ranges separated by comparisons of zonal mean monthly mean data of all profiles during the common observation period (Fig. 2) and collocated profile data (Fig. 3). The differences between the two types of comparisons are not significant.

The mean relative difference profiles between SCIAMACHY ozone and other satellites are shown in Figs. 4 and 5 for southern middle (30° S–60° S), tropical (30° S–30° N), and northern middle latitudes (30° N–60° N) for both direct zonal mean comparisons as well as for the collocation profile comparisons.

Generally, the agreement between both methods is very good, even in the case of comparison with solar occultation measurements (SAGE II, HALOE, ACE-FTS) which...
have a much lower sampling rate than the other limb sounders. This is consistent with the notion by Terao and Logan (2007) that the lower sampling rate of occultation instruments does not affect zonal mean calculation significantly. The computational costs for the analysis based on monthly means are much lower than for the collocation analysis, thus the monthly mean zonal mean comparisons are appropriate as indicated by the results in Table 1 and Figs. 2–5.

For most part of the stratosphere the mean agreement between SCIAMACHY and solar occultation data from HALOE and SAGE II is 10% and often better (≈5%) from 20 to 50 km altitudes. From the collocation comparisons, in particular, it is evident that HALOE has a small negative bias of about 5% with respect to SAGE II (collocated data), which is consistent with results from earlier studies (Nazaryan et al., 2005; Terao and Logan, 2007). SCIAMACHY is mostly in better agreement with SAGE II (few percent) than with HALOE. Up to 30 km altitude the comparison of SCIAMACHY with ACE-FTS is very similar to the results from the SAGE II comparisons, however, between 30 and 40 km SCIAMACHY is lower by 5% and above 40 km lower by 10%. The positive bias of upper stratospheric ACE-FTS ozone observed here is similar to that reported by Dupuy et al. (2009) with respect to various other data sets.

Between 20 and 40 km very good agreement between SCIAMACHY and MLS is observed (2–5%), with SCIAMACHY ozone concentrations being slightly higher than MLS above 25 km and slightly lower below. Above 40 km altitude SCIAMACHY is generally higher by 10%. Our results are consistent with the notion that MLS generally agrees to better than 10% with other correlative data with a tendency of a small negative bias in the upper stratosphere (Jiang et al., 2007; Froidevaux et al., 2008).

The agreement between SCIAMACHY and SABER is 10% and better above 30 km altitude in the tropics. In the upper stratosphere SCIAMACHY shows a similar negative bias as seen in the ACE-FTS comparison. At middle latitudes SCIAMACHY is lower than SABER by up to 20% near 35 km and lower by 10% and less above this altitude. For most part of the stratosphere SCIAMACHY is higher than SABER. The differences observed between SCIAMACHY and SABER cannot be explained by the day- and
nighttime differences since diurnal ozone variations are very small below 50 km altitude and at most a few percent.

A very common feature in the comparisons between SCIAMACHY and all other satellite data is the high bias below 20 km, particularly in the tropics, where differences can exceed 100%. The largest error source in the SCIAMACHY retrieval in this altitude region comes from high convective clouds obscuring the line of-sight (Sonkaew et al., 2009).

### 3.3 Significance analysis

To judge upon whether an observed difference between SCIAMACHY ozone profiles and other satellite data is statistically significant, an analysis in the framework of hypothesis testing is needed. Rodgers and Connor (2003) suggest to perform profile intercomparison using a χ² test. von Clarmann (2006) picked up the idea of comparing profiles $\hat{x}_{\text{val}}$ and $\hat{x}_{\text{ref}}$ from different instruments, where “val” denotes the instrument to be validated and “ref” represents the reference instrument. In this sense, the null-hypothesis is given by

$$H_0 : \hat{x}_{\text{val}}_h - \hat{x}_{\text{ref}}_h = 0, \quad \text{for all } h,$$

which means that measurements at each altitude $h$ of a given profile pair are equal, whereas the alternative is

$$H_1 : \hat{x}_{\text{val}}_h - \hat{x}_{\text{ref}}_h \neq 0, \quad \text{for at least one } h,$$

which is true, if measurements within a profile pair disagree for at least one altitude $h$. This approach makes sense if one wants to know if two profiles differ significantly or not by exploiting the full error budget, comprised of systematic, random and smoothing errors as given by the total error covariance matrix $S_{\text{total}}$ for each instrument (von Clarmann, 2006).

The quality of measurements from different instruments, however, depends on altitude, so the hypothesis testing can be applied for a certain altitude range for which
Eq. (6) has to hold. Another difficulty for the hypothesis testing is the lack of availability of the full error covariance matrix. To overcome this we modify the hypothesis testing from von Clarmann (2006) by using only the random error in each \((\ell, h, m)\) bin. This means that any significant difference between measurements has to be understood as a systematic bias excluding deviations by chance. However, the appeal of the \(\chi^2\) test also known as goodness-of-fit tests (e.g. Taylor, 1996; Fahrmeir et al., 2004) is its application to measurement vectors (not only scalars).

The standard deviation for each \((\ell, h, m)\) bin is defined as follows:

\[
\sigma_{U_{\ell,h,m}} = \sqrt{\frac{1}{K-1} \sum_{i=1}^{K} (u_i - U)^2}. \tag{8}
\]

The null-hypothesis can then be set up as follows:

\[
H_0: U_{\ell,h,m} - V_{\ell,h,m} = 0, \quad \text{for all } m, \tag{9}
\]

that the pairs of differences of mean profiles between two instruments for each zonal band \((\ell)\) and height \((h)\) bin are equal for all months \((m)\). The alternative is then

\[
H_1: U_{\ell,h,m} - V_{\ell,h,m} \neq 0, \quad \text{for at least one } m. \tag{10}
\]

The test statistic is then given by:

\[
\chi^2_0 = \sum_m \frac{(U_{\ell,h,m} - V_{\ell,h,m})^2}{\sigma^2_{d_{\ell,h,m}}}, \tag{11}
\]

where the standard deviation of the measurement differences is estimated by the squared sum of random errors from the individual mean profiles, i.e.

\[
\sigma_{d_{\ell,h,m}} = \sqrt{\sigma^2_{U_{\ell,h,m}} + \sigma^2_{V_{\ell,h,m}}}, \tag{12}
\]
to achieve independent and identically distributed (iid) $N(0,1)$ random variables in Eq. (11). The probability of observing a $\chi^2 \geq \chi^2_0$, i.e. $P(\chi^2 \geq \chi^2_0 | H_0)$ can be estimated by integrating a $\chi^2$ distribution with $m$ degrees of freedom from the actual value of $\chi^2$ to infinity. The null-hypothesis is typically rejected if $P(\chi^2 \geq \chi^2_0 | H_0) < 0.05$. Figure 6 shows the results from the significance test, where the rejection of $H_0$ is recommended for purple colored regions and red and gray denote regions, where the acceptance of $H_0$ is favored. The reason why very few intermediate probabilities between 0.05 and 0.95 are found is the relatively large number of summands in Eq. (11) and $m$ (degrees of freedom), respectively, which yields a rather small dynamical range of the $\chi^2$ cumulative distribution function. Areas of low probabilities (purple) in Fig. 6 elucidate regions, where the random uncertainties of measurements, which include also geographical and time mismatch (atmospheric variability), cannot be responsible for the observed differences in the mean profiles.

The hypothesis testing confirms the systematic and significant bias of SCIAMACHY in the tropics ($\pm 30^\circ$) below 20 km altitude with respect to most of the other instruments. From Fig. 2 it is evident that the differences in this region are systematically positive, meaning SCIAMACHY ozone overestimates ozone in the tropical UTLS region. A statistically significant bias is also seen for differences between SCIAMACHY and the other instruments MLS, SAGE II, and HALOE at around 30 km and at 40 km altitude (only 30 km for SABER) in the tropics where the biases approach up to about $+10\%$ (Figs. 2 and 3). In the extratropics the biases are mostly insignificant, which may be due to the enhanced natural variability as compared to the tropics. In the upper stratosphere the positive bias of SCIAMACHY with respect to HALOE and MLS is statistically significant.

Generally the patterns for the comparisons of SCIAMACHY appear similar for all instruments (Figs. 4, 5 and 6). Particularly in the tropics but to a lesser extent in the extratropics, the shape of the difference profiles shows a maximum near 30 and 40 km altitudes with biases of $+5$ to $+10\%$ and biases closer to zero and slightly negative at 25 km and near 35 km (statistically non-significant). The exact cause for this
is not known. Below about 35 km, most of the ozone information comes from the visible Chappuis absorption bands, while at higher altitudes contributions from the UV Hartley band absorption dominate. Inconsistencies in ozone absorption cross-section between Hartley and Chappuis as well between low and strong absorption regions may contribute to the variation in the biases as a function of altitude. Preliminary studies from a new SCIAMACHY retrieval version (V2.3) seem to indicate that inclusion of an aerosol climatology in the limb retrieval will most likely lower the bias below 35 km.

4 Trends

The estimation of trends in time series, mostly using least square regression, is a standard tool in climate research. However, the analysis of the significance of an observed trend is often ignored. The significance of a trend depends on the length of the time series, the noise, the autocorrelation of the noise, and also changes in instruments or measuring sites (Weatherhead et al., 1998). Atmospheric parameters like temperatures and ozone are often autocorrelated, e.g. if ozone is higher than normal in one specific month, it is likely, that ozone is also higher than normal the next months (Fioletov and Shepherd, 2003; Tegtmeier et al., 2008; Weber et al., 2011). Accounting for autocorrelations in a regression analysis does not generally change trend estimates but increase (for positive autocorrelations) the uncertainties of the trend and other fit parameters because of the reduction in degrees of freedom (e.g. Dhomse et al., 2006).

A crucial point in regression analysis is also the formulation of the trend model and its complexity. In the case of monthly mean ozone data the trend model should, additionally to the constant term and the linear trend, include harmonic terms representing seasonal variations and QBO (quasi-biennial oscillation) frequencies and phases. For instance Fischer and Tung (2008) revealed dominant QBO frequencies during the first decade in this century with periods of 26, 27, 28 and 29 months. Fitting all four nearby frequencies simultaneously may result in undesirable beat frequencies. The selection of terms to be included in a trend model is closely related to the problem of overfitting.
A concept to avoid this is known as *Occam’s razor*, which is the principle of choosing the hypothesis or model involving the fewest assumptions (and terms), if several models of different complexity describe phenomena equally well.

The SCIAMACHY time series between August 2002 and December 2008 contains a maximum of 47 monthly mean values, hence it is questionable to use the full model, including a constant, linear trend term, the annual and semiannual terms including phases and QBO terms (frequencies and phases). The selection of appropriate numbers of terms can be solved by a model selection criterion, e.g. the Bayesian Information Criterion (BIC) after Schwarz (1978):

\[
\text{BIC} = \ln(\sigma^2) + \frac{M}{T} \cdot \ln(T).
\] (13)

The BIC is an increasing function of the estimator of the variance \(\sigma^2\) of the fit residuals and the number \(M\) of fit parameters, while it decreases with the length \(T\) of the time series. The minimum of BIC provides the optimal compromise between small residuals and few fit parameters. We have calculated the BIC for trend models without any harmonic functions, with annual and semiannual components and up to four QBO frequencies for SCIAMACHY monthly mean data on the \((\ell, h)\) grid. For most time series at the \((\ell, h)\) grid we have found similar small BIC values for the model with annual and semiannual harmonic functions, and for the model including one QBO frequency. Hence, for reasons of consistent data processing we will use the trend model including one QBO period, which was chosen to be 28 months. A solar term accounting for the 11-yr solar cycle commonly applied in ozone trends is omitted here since the time periods considered here (maximum of six years) are too short. The six year period (2002–2008) falls into the descending phase of solar cycle 23 and solar cycle changes are contributing most likely to the derived linear trends. The trend model can be written as:

\[
Y_t = \mu + \omega X_t + S_t + N_t, \quad t = 0, ..., T.
\] (14)

The notation and handling of autocorrelation is described by Weatherhead et al. (1998)
and Mieruch et al. (2008). $Y_t$ are the monthly mean ozone data, $\mu$ is a constant, $\omega$, the linear trend, and $X_t$, the time with months as a running index (subscript $t$).

$S_t$ in Eq. (14) is the seasonal component including annual, semi-annual, and the QBO term, which is described as a Fourier series:

$$S_t = \sum_{j=1}^{2} [\beta_{1,j} \cdot \sin(2\pi j t/12) + \beta_{2,j} \cdot \cos(2\pi j t/12)] + [\beta_{1,3} \cdot \sin(2\pi t/28) + \beta_{2,3} \cdot \cos(2\pi t/28)].$$

(15)

$N_t$ in Eq. (14) contains the unexplained portion of the data, i.e. the noise, which is assumed to be an autoregressive process of the order of one, AR(1) (Schlittgen and Streitberg, 1997), i.e.

$$N_t = \phi N_{t-1} + \epsilon_t,$$

(16)

where $\epsilon_t$ are independent random variables with zero-mean and variance $\sigma^2\epsilon$. The magnitude or the memory of the autocorrelation is presented by $\phi$, which is restricted to $-1 < \phi < 1$, so the noise process $N_t$ is stationary. The memory of the data at lag one can be calculated using the autocorrelation function $\phi = \text{Corr}_{N_t,N_{t-1}}$, which is directly linked to the well known correlation coefficient. It has to be noted that accounting for an AR(1) process also considers autocorrelations at lags larger than one month, because the autocorrelation function of an AR(1) process decreases exponentially.

Generally the autocorrelation function is restricted to continuous, statistically stationary stochastic functions, or in the discrete case to equidistantly sampled data. Since there are gaps in our time series the discrete correlation function for analyzing unevenly sampled data was applied (Edelson and Krolik, 1988). Figure 7 shows the autocorrelations $\phi$ at lag $\tau = 1$ (one month) for SCIAMACHY ozone profiles during the period 2002–2008 for all zonal bands and altitudes. Maximum lag-1 auto-correlations are found in the tropical stratosphere.

After solving Eq. (14) in a least square sense we derive estimators for the fit parameters $\hat{\mu}$, $\hat{\omega}$, $\hat{\beta}_{ij}$ and their standard errors $\hat{\sigma}_{\mu}$, $\hat{\sigma}_{\omega}$, $\hat{\sigma}_{\beta_{ij}}$. A trend is statistically significant
at the 95% confidence level if it is greater than two times its error, i.e. $|\hat{\omega}| > 2\hat{\sigma}_\omega$ (Weatherhead et al., 1998). This follows from a z-test, which can be applied if the sample size is larger than 30 (Fahrmeir et al., 2004). Thus, the null-hypothesis would be that the trends ($\hat{\omega}$) are equal to zero $H_0 : \omega_0 = 0$.

One has to estimate the probability of observing an $\hat{\omega} \geq \omega_0$ given the null-hypothesis, i.e. $P(\hat{\omega} \geq \omega_0 | H_0)$, by integrating a zero mean Gaussian probability density function with standard deviation $\hat{\sigma}_\omega$ from $|\hat{\omega}|$ to infinity and additionally multiplying this probability by a factor of two to account for a two tailed test, because no prior information is given on the sign of the trend. Thus one finds that $P(\hat{\omega} \geq \omega_0 | H_0) < 0.05$ if $|\hat{\omega}| > 2\hat{\sigma}_\omega$, then suggesting rejecting the null-hypothesis.

As an example we show time series of SCIAMACHY and MLS monthly mean ozone observations in the zonal band $0^\circ$–$10^\circ$ S at 40 km altitude, respectively, in Fig. 8. The figure depicts SCIAMACHY (orange) and MLS (green) monthly mean data along with the linear trend (dark grey) and the fitted trend model from Eq. (14). Also shown in this figure are the SCIAMACHY trends for the shorter period (2004–2008) coinciding with the MLS observation period, which allows a direct comparison of trends. The linear trend in SCIAMACHY ozone is statistically significant for the period 2002–2008, but insignificant for MLS and SCIAMACHY for the shorter time span 2004–2008.

In Fig. 9 linear trends of SCIAMACHY limb ozone are shown for each latitude-altitude bin for the different time periods of overlaps with correlative satellite data, 2002–2008 (SABER), 2002–2005 (HALOE, SAGE II), and 2004–2008 (MLS, ACE-FTS). Positive and negative trends are shown in separate panels. Statistically significant negative trends (95% confidence level) are found in the 35–40 km altitude range varying from $-0.5\%$ yr$^{-1}$ in the northern hemisphere to $-1\%$ yr$^{-1}$ in the tropics and southern hemisphere from 2002 to 2008. From Fig. 9 it is evident that the negative trends are stronger in the early period from 2002 to 2005 as compared to the later 2004–2008 period. Statistically significant negative trends are seen below 20 km in the tropics and northern latitudes. Above this altitude trends are positive but mostly insignificant at the 95% confidence level. In addition to the significance criterion from above (trend larger than...
two times its error), we extend our significance criterion by requiring that missing data make up less than 10% of the number of possible data in the timeseries.

The SCIAMACHY trends shown in Fig. 9 can be compared to the trends derived for the correlative data as shown in Fig. 10 (positive trends) and Fig. 11 (negative trends). Very few statistically significant trends have been observed during the rather brief periods studied here. The trends from SAGE II, HALOE, and ACE-FTS appear more scattered due to the lower sampling of solar occultation instruments, however, they are mostly small and statistically non-significant. None of the correlative data confirm the statistically significant negative trends observed by SCIAMACHY. This clearly points to an instrumental issue with SCIAMACHY and the most likely explanation is the effect from a residual tangent height error in the SCIAMACHY limb measurements.

Unpublished results with a significantly improved version of the TRUE (Tangent height Retrieval by UV-B Exploitation) tangent height retrieval algorithm (Kaiser et al., 2003; von Savigny et al., 2005a) show evidence for a negative trend in the limb tangent height errors at tropical latitudes of about 50 m yr\(^{-1}\) from 2003 up to mid-2006, consistent with an anomalous trend in ozone concentrations between 30 and 40 km of about \(-1\) % yr\(^{-1}\) (von Savigny et al., 2009). More recent investigations showed that evidence for this anomalous tangent height trend is not seen any more after mid-2006. This behavior observed in the retrieved tangent height offsets is qualitatively consistent with the finding, that according to this study statistically significant negative trends in ozone occur between 30 and 40 km for the period 2002–2005, but not for the period 2004–2008. Investigations as to the cause of this tangent height anomaly are ongoing.

As an alternative to comparing individual trends from different satellites, we also looked at trends in the SCIAMACHY minus satellite difference time series. In the latter case all trends (of the differences) were statistically insignificant.
5 Summary and conclusions

We compared long term data sets of ozone profiles from SCIAMACHY (2002–2008) with other satellite data from MLS, SABER, SAGE II, ACE-FTS, and HALOE. The comparisons were done for monthly mean zonal mean (10° latitude bins) and altitudes from 10 to 50 km in steps of one km. Monthly mean zonal mean ozone profiles were calculated for all instruments and directly compared as averages over the common observation periods (see Figs. 1, 2, and 4).

Between 20 and 50 km the agreement between SCIAMACHY and the other satellite data is better than 10% and often below 5%. A very common feature in the comparisons between SCIAMACHY and all other satellite data is the high bias below 20 km, particularly in the tropics, where differences can exceed 100%. The largest error source in the SCIAMACHY retrieval in this altitude region comes from high convective clouds (Sonkaew et al., 2009).

MLS, SABER, and SCIAMACHY have a large sample size of several ten thousand profiles (~250 000, ~50 000, and ~40 000 profiles, respectively) in each zonal band over the entire time span, while SAGE, HALOE and ACE-FTS provide only about 400 to 1200 profiles over the entire time span in each zonal band. The question arises if a direct comparison of zonal means is appropriate for solar occultation data because of the rather low sampling statistics. The comparisons have, therefore, been also limited to monthly means of collocated profiles as shown in Figs. 3 and 5. The results are very similar to the direct comparisons of zonal mean data from all available data. This is consistent with the notion by Terao and Logan (2007) that the sampling of the solar occultation data is sufficiently representative for calculating monthly mean zonal means. For long term comparisons, the direct comparisons of zonal mean monthly mean data is generally appropriate for validation purposes.

The interpretation of results have been supported by an additional hypothesis test that identifies if observed differences between SCIAMACHY and other instruments exceed random variations due to noise and geographical/time mismatching (atmospheric
variability) thus revealing systematic biases between instruments (Fig. 6) most likely related to retrieval or instrumental issues. One striking feature that is at least statistically significant in the tropics is that the difference profiles between SCIAMACHY and the other correlative data has a double peak structure with a minimum near 37 km and maxima near 30 and 45 km altitude (Figs. 2–5). The information content for the SCIAMACHY ozone profile comes from the Hartley band absorption (UV) for altitudes above 35 km and from the Chappuis absorption band (visible) below that altitude.

When long term measurements from multiple instruments are to be used in trend assessments an identification of possible instrument drifts is important. We applied a classical trend model using seasonal, QBO, and linear trend terms in order to study potential differences in trends between SCIAMACHY and the other satellites.

SCIAMACHY shows negative linear trends in the 35 to 40 km altitude range and below 20 km (tropics and northern latitudes) that are not confirmed by the correlative satellite data (Figs. 9–11). This points to some remaining issues with SCIAMACHY. The most likely explanation is a possible trend in the limb tangent height registration on the order of several tens of meters per year, affecting the ozone limb retrieval (von Savigny et al., 2009). The trend comparisons with the satellites covering the later period starting in 2004 (ACE-FTS, MLS) do not reveal a significant trend while the earlier period 2002–2005 (HALOE, SAGE II) does. This is consistent with the leveling off in the tangent height mismatch after about 2006 (von Savigny et al., 2009). Studies are ongoing to further reduce this error.

In an upgrade of the SCIAMACHY retrieval version 2.3 an aerosol climatology (rather than a fixed aerosol extinction and scattering profile) will be used in the retrieval and improvements in the lower stratospheric SCIAMACHY ozone profiles are expected. It is also planned to extend the comparisons to other available satellite data like MIPAS, GOMOS, SMR, and OSIRIS.

SCIAMACHY provides a near decadal ozone profile data set that is very valuable and complementary to other available satellite data sets. This data set has been used for long-term trend assessments (Jones et al., 2009; Steinbrecht et al., 2009), polar
ozone losses (Sonkaew et al., 2011), and studies on 27-day solar rotation signal in ozone (Dikty et al., 2010b).

Acknowledgements. This work has been supported by a grant from BMWi under grant number 50EE0839 (ENVIVAL-LIFE). 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.

References


Dikty, S., Weber, M., von Savigny, C., Sonkaew, T., Rozanov, A., and Burrows, J. P.: Modula-


Nazaryan, H., McCormick, M. P., and M., R. I. J.: New studies of SAGE II and HALOE...


Terao, Y. and Logan, J. A.: Consistency of time series and trends of stratospheric ozone as seen by ozonesonde, SAGE II, HALOE, and SBUV(2), J. Geophys. Res., 112, D06310,
vapor product for the Stratospheric Aerosol and Gas Experiment (SAGE) II version 6.2 data

von Clarmann, T.: Validation of remotely sensed profiles of atmospheric state variables: strate-
gies and terminology, Atmos. Chem. Phys., 6, 4311–4320, doi:10.5194/acp-6-4311-2006,
2006. 4880, 4881

T.: Spatial and temporal characterization of SCIAMACHY limb pointing errors during the first
three years of the mission, Atmos. Chem. Phys., 5, 2593–2602, doi:10.5194/acp-5-2593-2005,
2005a. 4887

von Savigny, C., Rozanov, A., Bovensmann, H., Eichmann, K.-U., Noël, S., Rozanov, V. V.,

von Savigny, C., Bovensmann, H., Bramstedt, K., Dikty, S., Ebojie, F., Jones, A., Noël, S.,
in the SCIAMACHY Level 1 version 6.03 tangent height information, SCIAMACHY technical
note TN-IUP-scia-pointing-2009-01, available upon request from csavigny@iup.physik.
uni-bremen.de, 2009. 4887, 4889

Assessment of SAGE version 6.1 ozone data quality, J. Geophys. Res., 107, 4691,

Waters, J., Froidevaux, L., Harwood, R., Jarnot, R., Pickett, H., Read, W., Siegel, P., Cofield, R.,
Filipiak, M., Flower, D., Holden, J., Lau, G., Livesey, N., Manney, G., Pumphrey, H., Santee,
M., Wu, D., Cuddy, D., Lay, R., Loo, M., Perun, V., Schwartz, M., Stek, P., Thurstans, R.,
Boyles, M., Chandra, K., Chavez, M., Chen, G.-S., Chudasama, B., Dodge, R., Fuller, R., Gir-
rard, M., Jiang, J., Jiang, Y., Knosp, B., LaBelle, R., Lam, J., Lee, K., Miller, D., Oswald, J., Pat-
tel, N., Pukala, D., Quintero, O., Scaife, D., Van Snyder, W., Tope, M., Wagner, P., and Walch,
M.: The Earth observing system Microwave Limb Sounder (EOS MLS) on the Aura Satellite,
4875

Weatherhead, E. C., Reinsel, G. C., Tiao, G. C., Meng, X.-L., Choi, D., Cheang, W.-K., Keller,
T., DeLuisi, J., Wuebbles, D. J., Kerr, J. B., Miller, A. J., Oltmans, S. J., and Frederick,
**Table 1.** Summary of global (60° S–60° N) mean percent differences between SCIAMACHY and other satellite data. The difference profiles for zonal mean comparisons and for collocated profile data are shown in Figs. 2 and 3, respectively.

<table>
<thead>
<tr>
<th></th>
<th>10–20 km</th>
<th>20–30 km</th>
<th>30–40 km</th>
<th>40–50 km</th>
</tr>
</thead>
<tbody>
<tr>
<td>SCIA-MLS (zonal mean)</td>
<td>15.9±13.2</td>
<td>5.4±4.2</td>
<td>3.0±3.2</td>
<td>10.0±3.5</td>
</tr>
<tr>
<td>SCIA-MLS (collocated)</td>
<td>9.7±14.3</td>
<td>2.8±5.2</td>
<td>4.2±2.7</td>
<td>9.5±4.1</td>
</tr>
<tr>
<td>SCIA-SABER (zonal mean)</td>
<td>–</td>
<td>–</td>
<td>−7.4±4.0</td>
<td>−7.9±4.8</td>
</tr>
<tr>
<td>SCIA-SABER (collocated)</td>
<td>–</td>
<td>–</td>
<td>−4.8±3.3</td>
<td>−6.6±4.4</td>
</tr>
<tr>
<td>SCIA-SAGEII (zonal mean)</td>
<td>26.8±12.4</td>
<td>2.8±1.4</td>
<td>0.7±2.8</td>
<td>0.2±3.3</td>
</tr>
<tr>
<td>SCIA-SAGEII (collocated)</td>
<td>9.7±12.5</td>
<td>−0.3±2.8</td>
<td>−0.1±1.8</td>
<td>3.3±3.3</td>
</tr>
<tr>
<td>SCIA-HALOE (zonal mean)</td>
<td>52.1±23.8</td>
<td>5.7±2.1</td>
<td>1.0±3.1</td>
<td>7.4±2.5</td>
</tr>
<tr>
<td>SCIA-HALOE (collocated)</td>
<td>64.6±38.8</td>
<td>6.4±1.3</td>
<td>4.9±1.6</td>
<td>10.3±1.8</td>
</tr>
<tr>
<td>SCIA-ACEFTS (zonal mean)</td>
<td>26.5±13.3</td>
<td>2.6±1.6</td>
<td>−5.7±3.8</td>
<td>−9.9±7.1</td>
</tr>
<tr>
<td>SCIA-ACEFTS (collocated)</td>
<td>12.9±13.0</td>
<td>−1.0±3.3</td>
<td>−4.0±2.6</td>
<td>−9.9±6.7</td>
</tr>
</tbody>
</table>
Fig. 1. Top: monthly mean zonal mean ozone profiles from SCIAMACHY, MLS, SABER, SAGE II, HALOE, and ACE-FTS (August 2004, 20° S–30° S). Bottom: percent differences between SCIAMACHY and the other satellite instruments.
Fig. 2. Mean percent differences in monthly mean zonal means between SCIAMACHY and the other instruments, MLS, SABER, SAGE II, HALOE and ACE-FTS.
Fig. 3. Mean percent differences between SCIAMACHY and the other instruments, MLS, SABER, SAGE II, HALOE and ACE-FTS based on monthly means of collocated profiles. Collocation radius of 400 km and maximum time difference of 4 h, all instruments, and 24 h, SABER only.
Fig. 4. From left to right: percent differences in zonal means between SCIAMACHY ozone and other satellite data in zonal bands of 30° S–60° S, 30° S–30° N and 30° N–60° N.
Fig. 5. Same as Fig. 4 but for zonal means only derived from collocated profiles.
Fig. 6. $\chi^2$ test on differences between SCIAMACHY monthly mean zonal mean ozone profiles and MLS, SABER, SAGE, HALOE and ACE-FTS data. Purple patterns indicate significant differences which cannot be explained by random errors.
Fig. 7. Autocorrelation $\phi$ at lag $\tau = 1$ (one month) for SCIAMACHY ozone profiles from 2002 to 2008.
Fig. 8. SCIAMACHY (orange) and MLS (green) monthly mean ozone timeseries near the equator (−10° to 0°) at 40 km altitude. The black curves represent the complete fit (QBO and linear trend) and the lines depict the linear trends. The linear trend of SCIAMACHY ozone over the longer period 2002–2008 (top panel) is statistically significant (2σ), while both MLS and SCIAMACHY trends for the shorter common observation period 2004–2008 are insignificant (bottom two panels).
Fig. 9. Linear trends of SCIAMACHY ozone in percent per year for three time periods: 2002–2008 (top), 2004–2008 (middle), and 2002–2005 (bottom). Left panels show the positive trends, right panels the negative trends. Thick black contour indicate statistical significance at the 95% confidence level.
Fig. 10. Positive trends $\% \text{yr}^{-1}$ from MLS, SABER, SAGE II, HALOE and ACE-FTS. Thick black contours indicate significant trends with $\omega > 2\sigma$. 
Fig. 11. Same as Fig. 10 but showing only the region of negative linear trends.