Performance of the ground-based total ozone network assessed using satellite data

V. E. Fioletov,[1] G. Labow,[2] R. Evans,[3] E. W. Hare,[1] U. Köhler,[4] C. T. McElroy,[1] K. Miyagawa,[5] A. Redondas,[6] V. Savastiouk,[1] A. M. Shalamyansky,[7] J. Staehelin,[8] K. Vanicek,[9] and M. Weber[10]

Received 10 January 2008; revised 14 March 2008; accepted 26 March 2008; published 31 July 2008.

[1] Dobson and Brewer spectrophotometer and filter ozonometer data available from the World Ozone and Ultraviolet Data Centre (WOUDC) were compared with satellite total ozone measurements from TOMS (onboard Nimbus 7, Meteor 3, and Earth Probe satellites), OMI (AURA satellite) and GOME (ERS-2 satellite) instruments. Five characteristics of the difference with satellite data were calculated for each site and instrument type: the mean difference, the standard deviation of daily differences, the standard deviation of monthly differences, the amplitude of the seasonal component of the difference, and the range of annual values. All these characteristics were calculated for five 5-year-long bins and for each site separately for direct sun (DS) and zenith sky (ZS) ozone measurements. The main percentiles were estimated for the five characteristics of the difference and then used to establish criteria for "suspect" or "outlier" sites for each characteristic. About 61% of Dobson, 46% of Brewer, and 28% of filter stations located between 60°S and 60°N have no "suspect" or "outlier" characteristics. In nearly 90% of all cases, Dobson and Brewer sites demonstrated 5-year mean differences with satellites to be within ±3% (for DS observations). The seasonal median difference between all Brewer DS measurements at 25°–60°N and GOME and OMI overpasses remained within ±0.5% over a period of more than 10 years. The satellite instrument performance was also analyzed to determine typical measurement uncertainties. It is demonstrated that systematic differences between the analyzed satellite instruments are typically within ±2% and very rarely are they outside the ±3% envelope. As the satellite instrument measurements appear to be better than ±3%, ground-based instruments with precision values worse than ±3% are not particularly useful for the analyses of long-term changes and comparison with numerical simulations.

Citation: Fioletov, V. E., et al. (2008), Performance of the ground-based total ozone network assessed using satellite data, *J. Geophys. Res.*, *113*, D14313, doi:10.1029/2008JD009809.

1. Introduction

[2] Detection of ozone decline and the expected future ozone recovery requires long-term records of ozone measurements with small uncertainties. The expected rate of ozone increase due to the reduction of ozone-depleting substances is about 1% per decade [e.g., *World Meteorological Organization* (*WMO*), 2007]. Even a small long-term drift in the measurements could yield a large error in the estimate of ozone recovery. For a merged satellite data set *Stolarski and Frith* [2006] estimated an overall drift uncertainty of a little more than 1%/decade due to instrument-related effects. Ground-based Dobson and Brewer spectrophotometers can potentially maintain a 1% level of precision over long time intervals [*Komhyr et al.*, 1989; *Basher*, 1982, 1994; *Kerr et al.*, 1997; *Fioletov et al.*, 2005]. Since instrument calibration errors are typically independent for individual ground-based instruments, ozone values averaged from several stations should have even smaller uncertainties. However, to operate on that level, instruments must be properly and regularly calibrated and maintained, and the data should be appropriately processed and, if necessary, reprocessed.

[1]Environment Canada, Toronto, Ontario, Canada.
[2]SSAI- Lanham and NASA Goddard Space Flight Center, Greenbelt, Maryland, USA.
[3]NOAA Earth System Research Laboratory/Global Monitoring Division, Boulder, Colorado, USA.
[4]Meteorological Observatory Hohenpeissenberg, German Weather Service, Hohenpeissenberg, Germany.
[5]Japan Meteorological Agency, Tsukuba, Japan.
[6]Izana Observatory, National Institute of Meteorology, Santa Cruz de Tenerife, Spain.
[7]Main Geophysical Observatory, St. Petersburg, Russia.
[8]Institute for Atmospheric and Climate Science, Zürich, Switzerland.
[9]Czech Hydrometeorological Institute, Hradec Kralove, Czech Republic.
[10]Institute of Environmental Physics and Remote Sensing, University of Bremen, Bremen, Germany.

Copyright 2008 by the American Geophysical Union.
0148-0227/08/2008JD009809

[3] Ground-based data are widely used for the validation of satellite ozone retrievals [e.g., *Fleig et al.*, 1986; *McPeters and Labow*, 1996; *Bramstedt et al.*, 2003; *Weber et al.*, 2005; *Balis et al.*, 2007a, 2007b]. Satellite data have been used to estimate the performance of the ground-based network as well to detect potential problems with individual station records [*Bojkov et al.*, 1988; *Fioletov et al.*, 1999]. Since a single version of the analysis algorithm is used to process the entire data set from a single satellite instrument and satellite instrument characteristics are typically changing slowly, sudden "jumps" or large drifts for individual sites in the ground-satellite differences are commonly related to problems with ground-based data. Comparison with satellite data was recommended as a diagnostic tool for Dobson data reevaluation [*WMO*, 1993]. There are, however, certain limits on how well ground-based and satellite data can agree. The limits are controlled by a difference in the algorithms and assumptions used to calculate total ozone, e.g., by the difference in ozone absorption coefficients, their temperature dependence, assumptions about the stratospheric ozone and temperature distribution, and radiative transfer calculations. Ground-based and satellite instruments also measure different characteristics of total ozone: while the former instruments measure column ozone along the path in the direction toward the Sun, the latter retrieve ozone over a large area of the instrument field-of-view. Sources of the difference between ground-based and satellite systems has been studied intensively [e.g., *Vanicek*, 2006; *Balis et al.*, 2007a].

[4] At present, an agreement better than 1% between ground-based and satellite data is likely not achievable, but many stations have standard deviations of monthly mean differences with satellite data less than 2%. In 1986–1993 more than 72% of all Dobson stations had standard deviations at or below this level [*Fioletov et al.*, 1999]. Large differences with satellite data can often be traced to problems in ground-based records. Detection of such problems and the selection of a set of "good" stations for such tasks as satellite measurement validation or estimation of long-term ozone trends is often done subjectively. In this study we focus on the distribution of various characteristics of the difference between satellite and ground-based data and the application of these characteristics for uncovering potential problems with the ground-based data records.

[5] Section 2 describes the data sets and comparison algorithm. General results of ground-based and satellite data comparisons for wide latitudinal belts are discussed in section 3.1. The uncertainties associated with satellite data were estimated based on a comparison of data from different satellite instruments. Results of this comparison for gridded and overpass satellite data are discussed in sections 3.2 and 3.3 respectively. The characteristics of the difference between ground-based and satellite data are given in section 3.4 and the overall performance of the ground-based network is reviewed in section 3.5. A summary and a discussion of the results are given in section 4. The performance of individual sites will not be discussed in this paper. A station-by-station report of the comparisons between the various satellites and individual sites statistics is available from the World Ozone and Ultraviolet Radiation Data Centre (WOUDC) web site (www.woudc.org).

2. Data Set and Comparison Algorithms

2.1. Satellite Data

[6] Data from the Total Ozone Mapping Spectrometer (TOMS) instrument onboard the Nimbus 7 (from November 1978 to May 1993), Meteor 3 (from August 1991 to November 1994) and Earth Probe (EP) (from July 1996 to December 2005) satellites and the Ozone Monitoring Instrument (OMI) onboard the Aura satellite (from August 2004 to December 2007) all processed using the version 8 algorithm (available from http://toms.gsfc.nasa.gov) were used in this study. Also used were data from the Global Ozone Monitoring Experiment (GOME) onboard the European Space Agency's ERS-2 satellite (from July 1995 to June 2003). Three algorithms were developed for GOME applications in 2003 and used to reprocess the GOME total ozone data. These algorithms are all based on the DOAS (Differential Optical Absorption Spectroscopy) technique. They are the WFDoas algorithm [*Coldewey-Egbers et al.*, 2005]; the TOGOMI/TOSOMI algorithm [*Eskes et al.*, 2005]; and the GDOAS algorithm [*Van Roozendael et al.*, 2006]. For all three algorithms, retrieved GOME total ozone show excellent agreement with each other, so only one GOME data set was used here (WFDoas version 1 algorithm). The GOME data are available from http://www.iup.uni-bremen.de/gome/wfdoas/.

[7] For all satellite data sources, the analysis was done using 1° latitude by 1.25° longitude gridded (Level 3) data with daily temporal resolution. Ozone values for individual sites (overpasses) were also extracted from the gridded data. Gridded data were preferred in order to avoid issues related to the differences in fields-of-view of the individual instruments so that satellite-to-satellite comparisons were more appropriate. In regions where there are multiple orbits over the same location, the gridded data product also has the advantage of being an average of measurements made at different times, more closely reproducing the average daily value measured from the ground.

[8] While a discussion about different satellite algorithms is beyond the scope of this paper, one principal difference between WFDoas and Version 8 TOMS algorithms should be mentioned. Both algorithms estimate "ghost" vertical column to account for missing ozone below clouds and thus add it to the measured column ozone. WFDoas algorithm uses effective scene height determined from cloud top height and cloud fraction information derived in the oxygen A-band in combination with the terrain height of the observed GOME ground pixel [*Coldewey-Egbers et al.*, 2005]. In the TOMS algorithm, reflectivity measurements and cloud climatology are used to determine cloud top height [*McPeters et al.*, 1998]. Therefore it is expected to see some difference between GOME and TOMS/OMI ozone retrieved under cloudy conditions.

[9] The difference between satellite and ground-based data show practically no dependence on the solar zenith angle and surface reflectivity for the most recent versions of satellite algorithms discussed here except for observations at very large zenith angles (greater that 80°) seen at high latitudes [*McPeters and Labow*, 1996; *Weber et al.*, 2005;

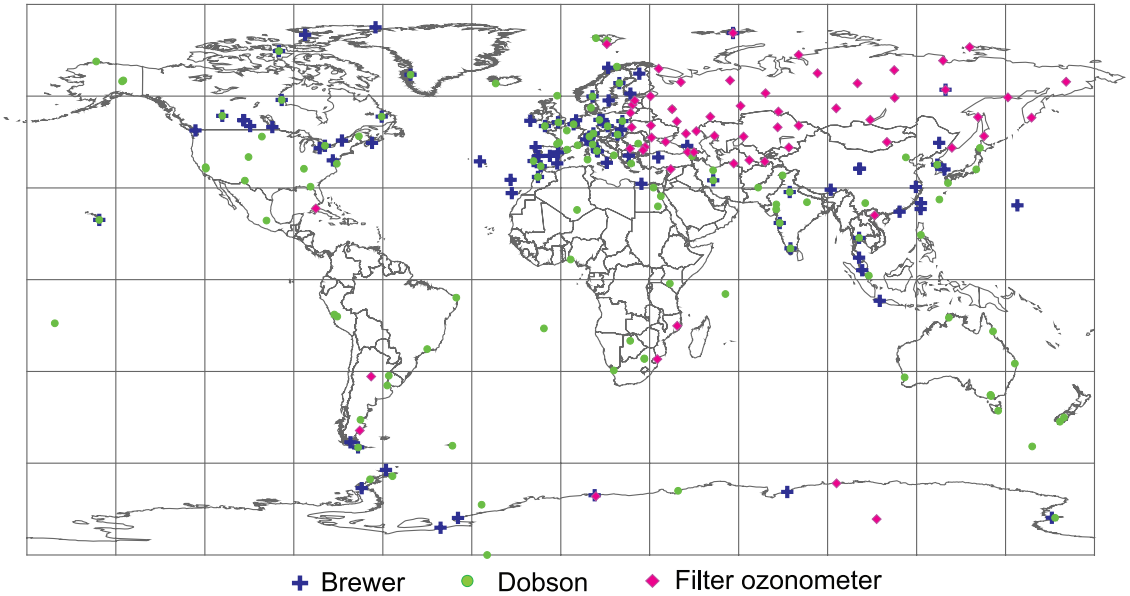

Figure 1. Map showing the location of total ozone stations where data were submitted to the WOUDC within the period of 1979–2006. The WOUDC data contained within this temporal range represents a total of 103 agencies, located in 78 countries, reporting data.

Balis et al., 2007b]. Since this study is focused on stations that are outside the polar regions, we did not put any criteria on the range of zenith angles for satellite observations.

2.2. Ground-Based Data

[10] Daily mean values derived from Dobson and Brewer spectrophotometer and filter ozonometer measurements are available from the WOUDC archive and have been analyzed in this study for the period 1978–2006. A map of the station locations is shown in Figure 1. Calculations are done separately for direct sun (WOUDC observation code is 0 or "DS") and zenith sky observations (WOUDC observation codes between 2 and 7 or "ZS"). Typically, a station reports a daily mean value (or "daily representative" value for some Dobson sites) from DS measurements if one or more DS observations are available, otherwise the daily mean value is calculated from ZS observations. In other words, ZS daily values are typically reported on cloudy days. Dobson DS measurements can be performed using different wavelength pairs (AD, CC′, and CD-pair, see *Komhyr* [1980] for details), but since 1978, more than 98% of Dobson DS daily values have been obtained using the AD-pair.

[11] The measurements of the ground-based network of the Dobson and Brewer network rely on the primary instruments. The primary Dobson instrument is maintained by NOAA at Boulder, USA [*Komhyr et al.*, 1989], while the Brewer network calibration is based on a triad of Brewer instruments operated at Toronto, Canada, by Environment Canada [*Kerr et al.*, 1997; *Fioletov et al.*, 2005]. The calibration scales of Dobson and Brewer networks are independent. The vast majority of filter ozonometer are calibrated against the Russian standard Dobson 108, which is part of the European Dobson calibration system, and data are processed by the Russian Main Geophysical Observatory (MGO), St. Petersburg.

[12] There are numerous publications about the instruments and their performance [e.g., *Komhyr*, 1980; *Basher*, 1982; *Gustin et al.*, 1985; *Kerr et al.*, 1988; *Bojkov et al.*, 1994] and therefore they are not discussed in any detail here. However, basic information on calibration procedures is given below to explain potential sources of errors in the ground-based data. All ground-based instruments are calibrated with two main parameters estimated: the extraterrestrial constant (ETC) value and the effective ozone absorption coefficient. A calibrated reference "standard" instrument is typically used to establish the ETC value. An effective ozone absorption coefficient can be derived directly from tests for the individual instrument, as is currently done for Brewer instruments, or the instrument effective ozone absorption coefficients can be taken as given and then the individual instrument aligned to the same specifications as the standard instrument, as is done for Dobson instruments [*Dobson*, 1957]. An error in the ETC value yields an error that depends on the solar zenith angle, while an error in effective ozone absorption coefficient introduces a relative bias. While it is strongly recommended that data be evaluated and, if needed, reprocessed after calibrations, this is not always done.

Table 1. Data Bins With Corresponding Time Intervals and Satellite Data Sources Used in This Study

Bin	Period	Satellite and Instrument
1	1978–1985	Nimbus 7 TOMS (Version 8)
2	1986–1990	Nimbus 7 TOMS (Version 8)
3	1991–1995	Nimbus 7 TOMS (Version 8), Meteor 3 TOMS (Version 8), GOME[a] (University of Bremen version)
4	1996–2000	EP TOMS (Version 8), GOME (University of Bremen version)
5	2001–2006	EP TOMS (Version 8), GOME, OMI (NASA version)

[a]GOME data are available starting in July 1995.

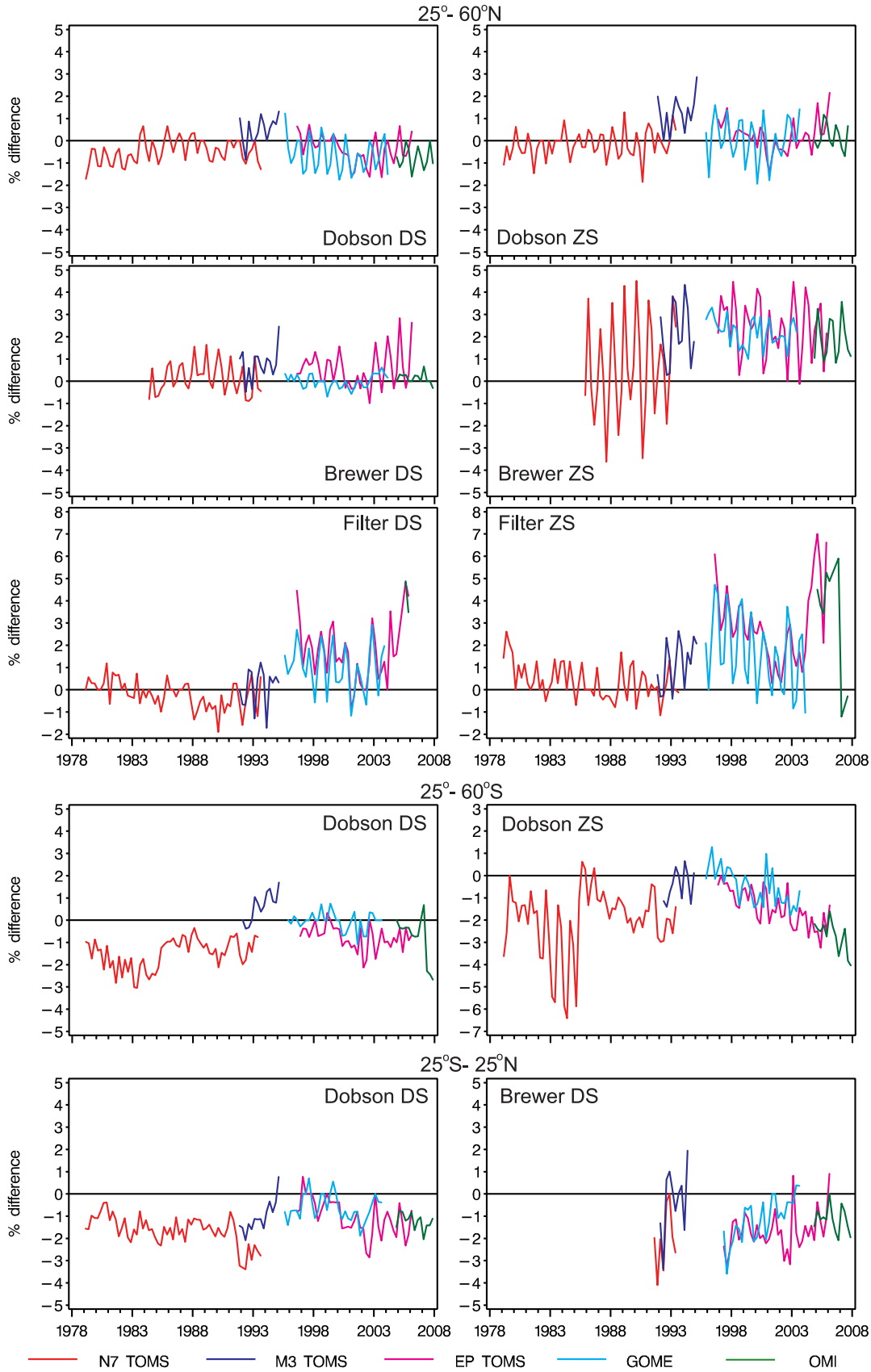

Figure 2. Difference between Dobson and Brewer daily total ozone values and data from different satellites over 25–60°N, 25–60°S, and the tropical region (25°S–25°N). Each point on the plot represents the median value of the difference for a season. Calculations were done for DS and ZS measurements separately.

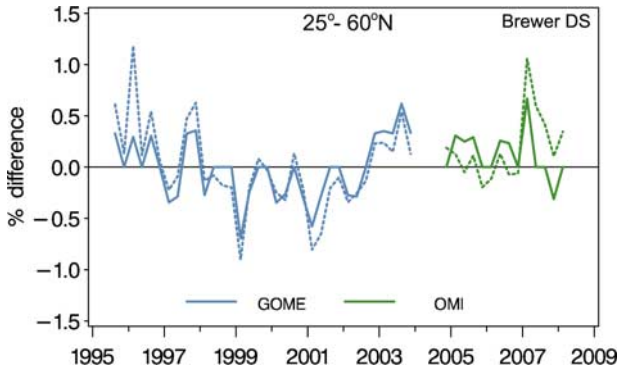

Figure 3. Mean (dotted line) and median (solid line) difference between Brewer DS and GOME (blue lines) and OMI (green lines) measurements at 25°–60°N. The plot possibly illustrated the best observed agreement between satellite and ground-based ozone observations. The median values are very similar to the mean values, but they are less sensitive to "outliers" and therefore demonstrate less scattering.

Without proper reprocessing, sudden "jumps" in the record of differences with the satellite data can be seen in some cases, corresponding to the recalibration at a specific date.

[13] There are small (typically within ±1.5%, but possibly larger in winter at high latitudes) seasonal, systematic differences between Dobson and Brewer total ozone [e.g., *Kerr et al.*, 1988; *Staehelin et al.*, 1998, 2003]. This difference in seasonal variation is partially attributable to the different temperature sensitivities of the ozone effective, differential absorption cross sections for the two types of instruments: unlike satellite algorithms, the operationally used Dobson and Brewer instrument total ozone retrieval algorithms do not account for temperature and ozone profile variability in the atmosphere [*Kerr*, 2002]. Different instrument stray light properties and Dobson instrument sensitivity to sulfur dioxide [*De Muer and De Backer*, 1992] are possible additional contributing factors to the Dobson-Brewer differences.

[14] The ground-based network is operated under the auspices of the Global Atmosphere Watch (GAW) program of the World Meteorological Organization (WMO) and Standard Operation Procedures (SOPs) are supposed to provide guidelines for observations. However, it should be noted that the stations shown in Figure 1 are operated by 103 different agencies, and operational procedures, calibration practices, data processing, etc., are not always exactly the same or conducted in a uniform manner by all the different agencies.

2.3. Comparison Algorithm

[15] All comparisons here were done using daily total ozone values with satellite overpasses matched with ground-based observations for each day and for every ground-based station. Then the difference between the ground-based and overpass values was calculated for each day at each site and the difference expressed as the percent of the average of ground-based and satellite values for that day. The obtained differences were then analyzed to determine the mean, the median and the standard deviation. For comparisons in section 3.1, daily percent differences from all sites within the analyzed latitudinal belt were lumped together to calculate seasonal differences. Calculations were done separately for each satellite instrument and for the DS and ZS observations of each of the three ground-based instrument types. The comparisons of different satellite data sets in section 3.2 were done using daily gridded data.

[16] Sections 3.3 and 3.4 discuss the distribution of certain characteristics of the difference among the sites. These characteristics were calculated for the difference between satellite data and overpasses (section 3.3) as well as between satellite data and ground-based observations (section 3.4). Because of changes in satellite instruments and fluctuations in time of ground-based instrument parameters, due primarily to instrument recalibrations and refurbishments, the analysis was done for five time intervals (or "bins") each of which was about 5-years long as shown in Table 1. Results were presented either for each bin separately or for estimates from all bins lumped together.

[17] Five characteristics of the differences were calculated for each site and instrument type: the mean difference, the standard deviation of daily differences, the standard deviation of monthly differences, the amplitude of the seasonal component of the difference, and the range of annual values. All of these characteristics were calculated for each bin and each site separately for DS and ZS measurements for the Dobson, Brewer, and filter instruments respectively.

[18] All percent difference values from one site within one bin were used to calculate the mean value and the standard deviation and at least 100 values were required. For the standard deviation of monthly differences, monthly mean differences were calculated for months with at least 7 days of data and then the standard deviation of these monthly mean differences was computed. At least 15 monthly values were required for this calculation. The range of annual values was calculated as the difference between the largest and the smallest annual mean difference. At least 60 days with data were necessary to be available to define an annual value of the difference. The limits on the number of values used here were set to be large enough to make the estimates of the parameters accurate and low enough to have a sufficient number of sites for the comparison.

[19] The amplitude of the seasonal difference was estimated for each station and each bin from the best fit of daily differences using the following function:

$$D(t) = \alpha + \gamma_1 \sin \omega t + \gamma_2 \cos \omega t + \text{residual}$$

where t is the time in months, $\omega = 2\pi/12$ months, α is an offset, and the $\gamma_1 \sin \omega t + \gamma_2 \cos \omega t$ term represents the seasonal component of the difference. The amplitude of the seasonal component was calculated as $\sqrt{\gamma_1^2 + \gamma_2^2}$. A minimum of least 300 days with data were required to be available for this calculation.

3. Results

3.1. Ground-Based Versus Satellite Data

[20] Figure 2 illustrates the overall performance of the ground-based network by presenting the percentage difference between Dobson, Brewer, and filter instruments and Nimbus 7, Meteor 3, Earth Probe TOMS, GOME, and OMI

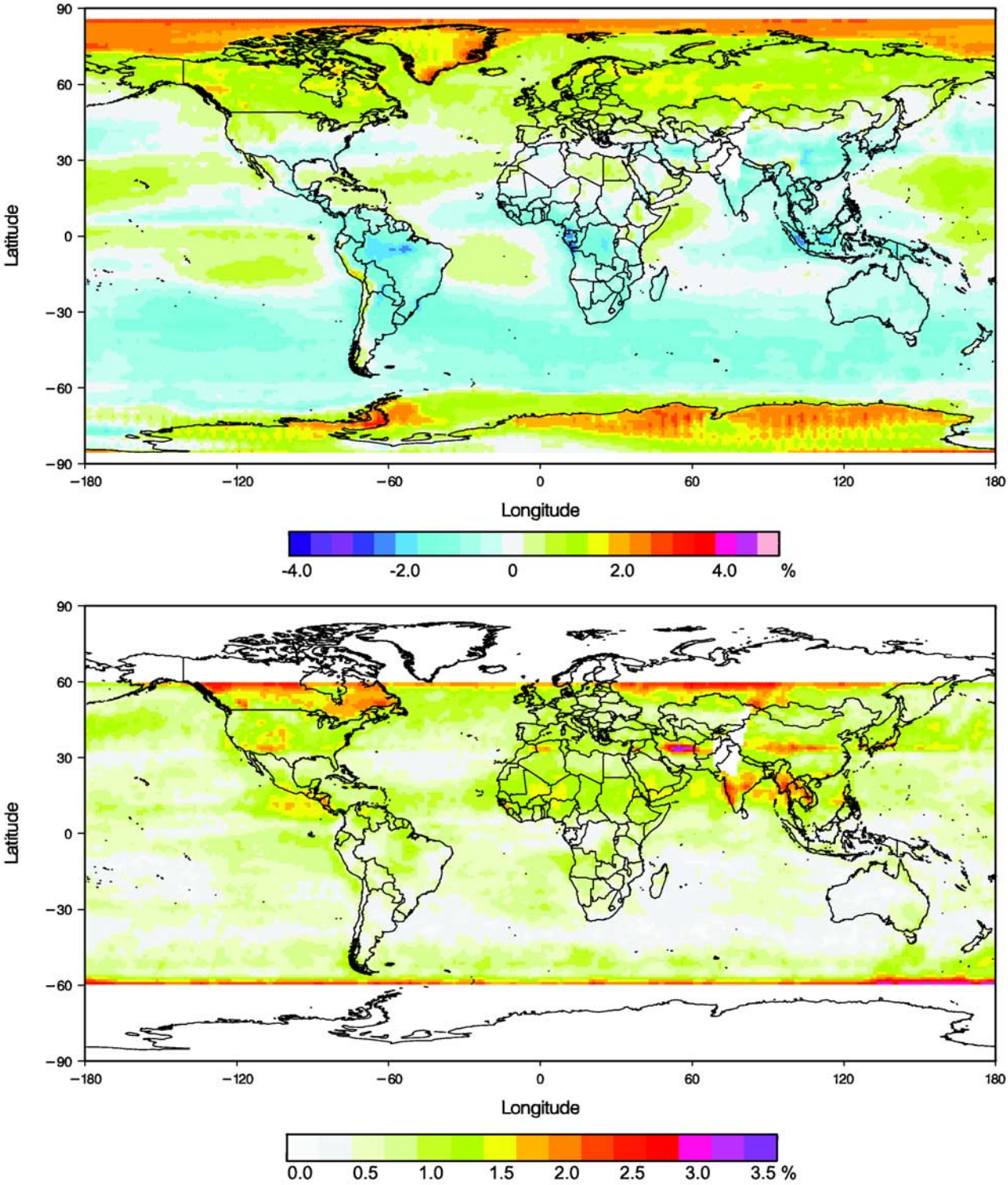

Figure 4. (top) The mean value of the GOME minus EP TOMS difference in percent for 1996–2000. (bottom) The seasonal amplitude of the GOME-EP TOMS difference in percent for the 60°S–60°N region for 1996–2000.

satellite instruments. Each point on the plot represents the median value of all differences between satellite overpass data and ground-based observations at all sites for one season. All available WOUDC data were used for these plots with no initial data screening. The median values were used instead of more common mean values to reduce the impact of "outliers" on the comparison results. Plots of the mean values are very similar with a slightly larger amplitude of the differences. Calculations are done separately for

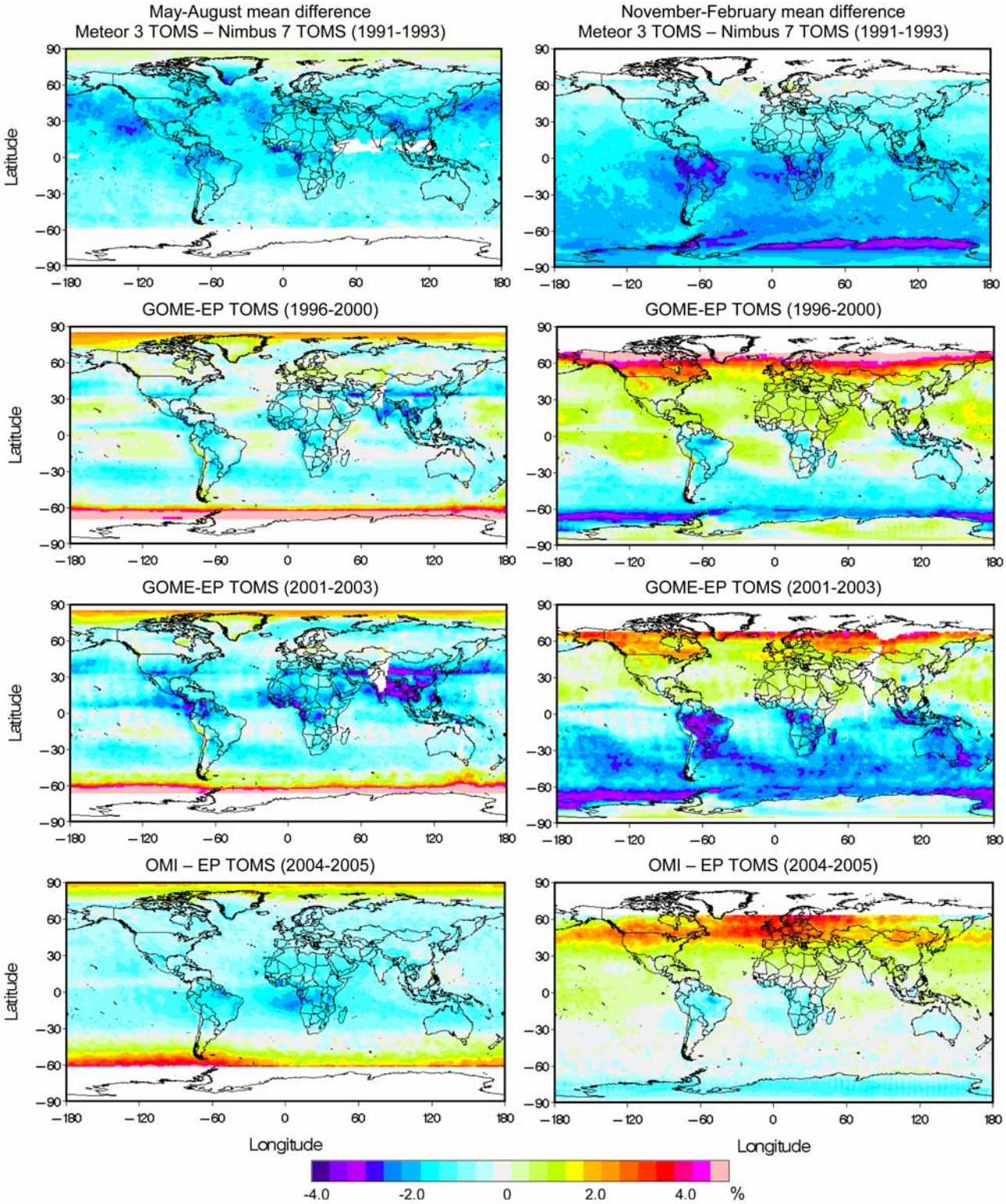

Figure 5. Mean differences between different satellite instruments estimated for May–August and November–February given in percent.

direct sun and zenith sky observations for three latitudinal belts (25°–60°N, 25°–60°S, and 25°S–25°N).

[21] The overall agreement between direct sun (DS) Dobson and Brewer observations and various satellite instruments at 25°–60°N is quite good and is close to the claimed instrument precision of about 1%. The agreement between Brewer DS data and GOME and OMI observations at 25°–60°N is truly remarkable with the median difference typically within ±0.5% range (Figure 3). Figure 2 also confirms a known result [*Weber et al.*, 2005] that in general Brewer DS observations show a better agreement with GOME and have some season-dependent difference with

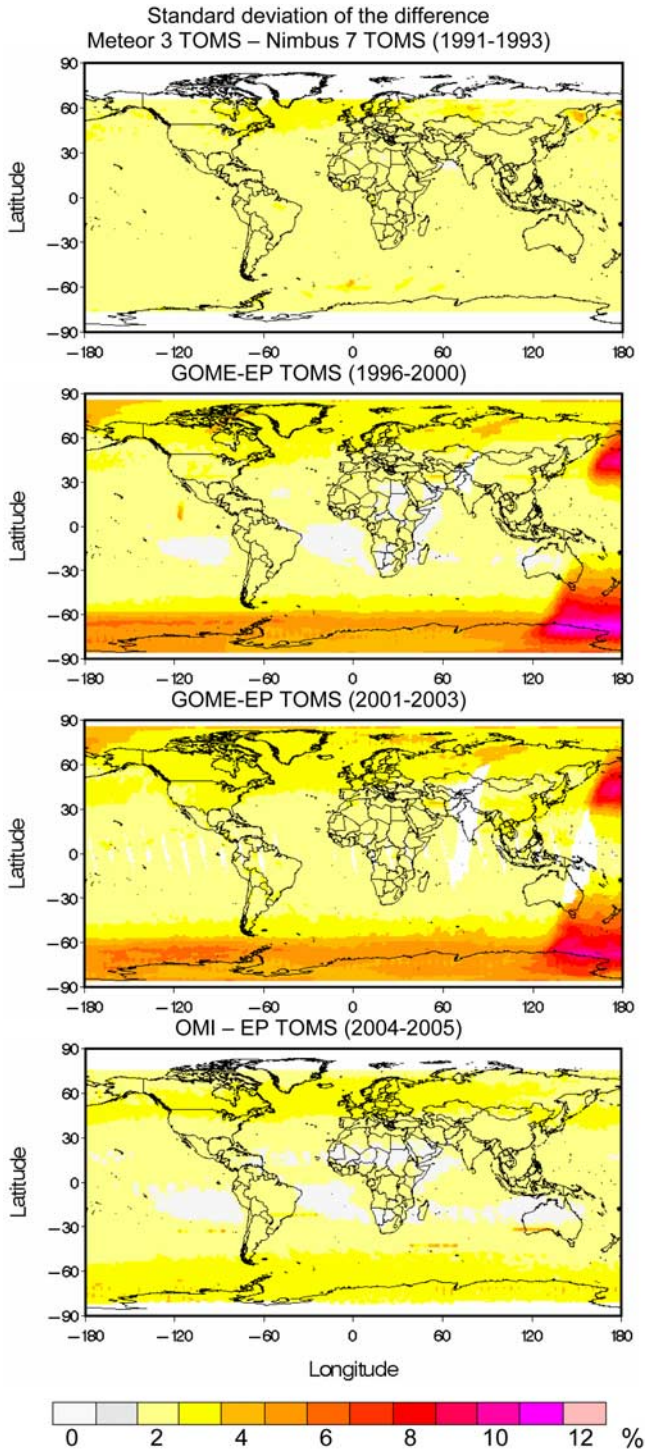

Figure 6. Standard deviation of the difference between the various satellite instruments for daily gridded data given in percent.

EP TOMS, while Dobson DS data have some season-dependent difference with GOME and not with EP TOMS. From Figure 2, there appears no such season-dependent difference between Dobson DS and EP data over the tropical belt and at 25°–60°S.

[22] The median difference of Dobson zenith sky (ZS) data with satellite observations is similar to those for Dobson DS data. Like the DS data, there is a clear seasonal dependence between Dobson ZS and GOME data while this dependence is much weaker when Dobson ZS data are compared with EP TOMS. The agreement between ZS daily values and satellite data for the 25°–60°S belt is not as good as for 25°–60°N belt. Brewer ZS data show a much larger seasonal dependence, particularly prior to 1995. They also appeared to be biased by ~2% thereafter. It should be mentioned, however, that only a small fraction (14%) of daily mean Brewer values were calculated from ZS observations. The same fraction for Dobson data is about 28%.

[23] The median difference of all filter ozonometer data against various satellite observations was typically within the ±1% range up to the mid-1990s and had a ~1.5% bias thereafter. Since the same bias was not seen in Dobson and Brewer records, this may suggest a systematic error of the entire filter ozonometer network. At the time of this study, no data after 2003 had been submitted to the WOUDC and the post-2003 differences are based on a very limited number of stations. Filter ozonometer ZS data demonstrate performance similar to filter ozonometer DS data.

[24] Figure 2 also shows a difference between individual satellite instruments: when data from two instruments are available at the same time, they can be compared to the same set of ground-based observations. This comparison demonstrates that the difference between satellite instruments is not just a uniform systematic bias. For example, Meteor 3 and Nimbus 7 TOMS data agree within ~1% at 25°–60°N, but there is a 1–2% bias between them over the tropical region (25°S–25°N) and in the southern hemisphere 25°–60°S (Nimbus 7 TOMS values are higher). The seasonal structure of the difference between Dobson and EP TOMS data seen at 25°–60°N, disappears over other two regions. The examples given above suggest that the difference between individual satellite instruments should be investigated before any conclusion about ground-based station performance is given based on satellite overpass data analysis.

3.2. Difference Between Individual Satellite Instruments

[25] Three pairs of satellite instruments were compared to study the mean, standard deviation, and the annual cycle amplitude of the differences. These three pairs represent three different scenarios: in the case of Meteor 3 and Nimbus 7 comparison, the same type of instrument and the same processing algorithm was used, but the satellite orbits were different. For the OMI-EP TOMS comparison, the same algorithm was used, but the instruments were significantly different. Finally, GOME and EP TOMS are different instruments and the data were processed using different algorithms. GOME and EP TOMS data were compared for two intervals, 1996–2000 and 2001–2003 to take into account serious degradation of the EP TOMS instrument after 2000. All comparisons were done using 1° latitude by 1.25° longitude gridded data.

[26] We start the discussion with the GOME and EP TOMS comparison for 1996–2000. Figure 4 (top) shows the mean difference (GOME minus EP TOMS expressed in % of the GOME and EP TOMS average). For about 90% of the grid cells, the mean difference is within the range of ±1.7%, there are areas where the difference is outside the

Table 2. Frequency Distribution for Various Parameters of the Difference Between Meteor 3, GOME, and OMI Overpasses With Nimbus 7 and EP TOMS Overpasses Calculated for WOUDC Sites[a]

Parameter	5th Percentile	10th Percentile	Median	90th Percentile	95th Percentile
DS Overpasses					
Mean difference, %	−1.5	−1.1	−0.2	0.8	1
Standard deviation of daily difference, %	1.4	1.5	1.9	2.5	2.6
Standard deviation of monthly difference, %	0.4	0.5	0.9	1.5	1.7
Seasonal amplitude of the difference, %	0.2	0.4	1.1	1.9	2
Range of annual mean differences, %	0.1	0.2	0.6	1.4	1.7
ZS Overpasses					
Mean difference, %	−2	−1.6	−0.2	1.3	1.8
Standard deviation of daily difference, %	1.5	1.6	2.3	3	3.2
Standard deviation of monthly difference, %	0.4	0.6	1.1	1.8	2
Seasonal amplitude of the difference, %	0.5	0.6	1.3	2.1	2.3
Range of annual mean differences, %	0.1	0.3	0.8	1.6	1.8

[a]Calculations were done separately for overpass data matching ground-based DS and ZS observations.

±2.5% interval. As discussed later, the standard deviation of the difference between the two daily gridded data sets is small. Therefore this high mean difference cannot be attributed to a random fluctuation.

[27] The areas of large differences are located over specific geographical regions and the isolines of the differences often follow the continental contours. These large differences are of particular interest in the tropical region where ozone variability is small and long-term fluctuations are typically within 1%. Over the tropical region, the difference ranges from +1.2% (i.e., GOME is higher then EP TOMS) over the Pacific and Atlantic to −2.8% over the Amazon River basin, Equatorial West Africa, and Sumatra. The interpretation of these differences is beyond the scope of this study, although we can point out a correlation between the difference distribution in the tropical region and region's cloud cover climatology.

[28] As Figure 2 suggests, the difference between the two satellite instruments may have an annual cycle. To estimate its amplitude, the difference time series for each grid point was fitted by a periodic function as described in section 2. The analysis was limited to the 60°S–60°N region to avoid issues related to missing data during the polar night. The results are shown in Figure 4 (bottom). For 90% of the grid cells located between 60°S and 60°N, the amplitude is less than 1.4%, however for 1% of all grid cells the amplitude is greater than 2.4%, particularly over Iran and the West coast of India, where the peak amplitude is as high as 3% (i.e., the minimum-to-maximum difference is 6%).

[29] The periods of parallel measurements of EP TOMS and OMI as well as Meteor 3 and Nimbus 7 TOMS are too short to study the annual cycle amplitude of the difference in the same way as it was done for GOME and EP TOMS pair. To illustrate the bias between the satellite instrument pairs and their dependence on season, Figure 5 shows the difference for the winter and summer seasons. Calculations are done for May–August and November–February periods that roughly represent ±2 month intervals around the summer and winter solstices. The plots of GOME-EP TOMS differences for 1996–2000 for winter and summer shows the features seen previously in Figure 4: a −3% bias in GOME against EP TOMS data over certain areas within the equatorial region, and a 3% positive bias can be seen over subpolar regions in winter time. For the period of 2001–2003, the GOME-EP TOMS difference shows essentially the same patterns as for 1996–2000, but with all values reduced by 0.5–1%. For example, if the November–February difference over the Amazon River basin was about −2.5% in 1996–2000, less than for the second period (2001–2003) when it was about −3.5%.

Table 3. Frequency Distribution for Various Parameters of the Difference Between Dobson, Brewer, and Filter Instrument Data With Satellite Overpasses[a]

Parameter	5th Percentile	10th Percentile	Median	90th Percentile	95th Percentile
Mean Difference, %					
Dobson DS	−3.6	−2.7	−0.6	1.1	1.5
Brewer DS	−2.7	−1.8	0.2	1.7	2.2
Filter DS	−1.8	−1.3	0.4	3.2	5.3
Dobson ZS	−4.8	−3.1	−0.2	1.8	2.3
Brewer ZS	−4.3	−1.7	2.5	6.5	7.4
Filter ZS	−1.9	−1.2	0.8	4.5	7.7
Standard Deviation of Daily Difference, %					
Dobson DS	1.7	1.9	2.5	4.2	5.5
Brewer DS	1.8	1.9	2.5	4.5	6
Filter DS	2.7	2.9	3.5	5.5	8.1
Dobson ZS	2.5	2.9	4.2	5.9	7.2
Brewer ZS	2.7	2.9	4.1	6.4	8.4
Filter ZS	3.7	3.9	4.7	6.8	7.6
Standard Deviation of Monthly Difference, %					
Dobson DS	0.8	0.9	1.3	2.3	3.3
Brewer DS	0.6	0.8	1.4	2.9	4.1
Filter DS	1.4	1.5	2	3.5	6.1
Dobson ZS	1	1.2	2	3.1	3.8
Brewer ZS	1	1.2	2.4	4.3	5.3
Filter ZS	1.6	1.8	2.6	4.1	5.1
Seasonal Amplitude of the Difference, %					
Dobson DS	0.16	0.2	0.7	1.5	1.9
Brewer DS	0.2	0.3	0.9	1.9	2.5
Filter DS	0.3	0.6	1.4	2.8	3.9
Dobson ZS	0.2	0.4	1.2	2.5	2.7
Brewer ZS	0.5	0.6	1.3	3.2	3.7
Filter ZS	0.4	0.5	1.4	3.1	3.9
Range of Annual Mean Differences, %					
Dobson DS	0.4	0.6	1.5	3.5	4.9
Brewer DS	0.2	0.5	1.4	4.5	6.5
Filter DS	0.5	0.8	2.2	5.3	9.4
Dobson ZS	0.5	0.7	1.8	4.6	5.3
Brewer ZS	0.2	0.4	1.8	3.9	6.5
Filter ZS	0.4	1.0	2.5	5.2	8.1

[a]Calculations were done separately for ground-based DS and ZS observations.

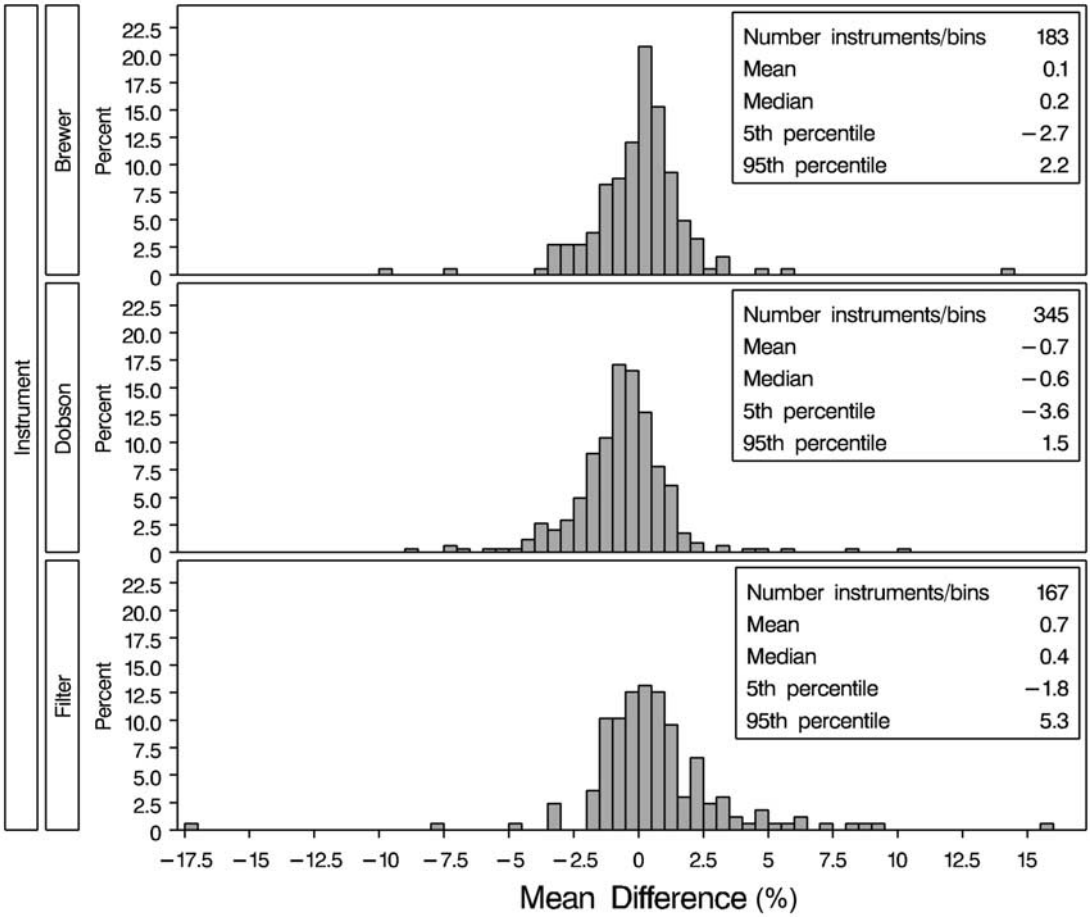

Figure 7. Histogram of distribution for the mean difference between ground-based DS observations and matched satellite overpasses for Dobson, Brewer, and filter instrument sites. The mean difference was estimated for each site and for each 5-year bin separately and then estimates from all bins were lumped together to produce the histogram.

[30] There is about a 16-month period when both OMI and EP TOMS data are available. It was at the end of EP TOMS instrument operation when the instrument experienced a severe degradation. Nevertheless, both winter and summer differences are typically within a ±2% range. OMI data were on average higher than EP TOMS by 2–3% at 40°–60°N, particularly over Europe and the Northern Atlantic but they were slightly (0.5–1%) lower than EP TOMS there in summer.

[31] The overlapping period of Nimbus 7 and Meteor 3 TOMS instruments was about 20 months but with some gaps due to the Meteor 3 satellite orbit. As Figure 5 shows, Meteor 3 data are typically lower than Nimbus 7 ozone values. The difference is small (less than 1%) north of ~45°N over the continents in both winter and summer, as was also seen in Figure 2. Over the tropical region and Southern Hemisphere, the deviations are negative in the two seasons shown in Figure 5. The difference there is particularly large, about 2–3%, in November–February. A greater than 3% difference can be seen over two spots in the tropics and over the Antarctica in November–February.

[32] Comparison of EP TOMS data with GOME in 2001–2003 and OMI in 2004–2005 shows that, over northern midlatitudes, the EP bias against GOME is nearly the same as the bias against OMI. If there was no drift in EP TOMS during 2001–2005, this may suggest that the difference between GOME and OMI is small over northern midlatitudes as was demonstrated earlier in Figure 3. However, this is not the case for southern midlatitudes and a 1–2% bias between GOME and OMI is expected there from comparisons with EP TOMS.

[33] The standard deviations of the difference between the various satellite instruments are shown in Figure 6. Standard deviations of the difference between daily values are typically 2–3% and are not much different for all satellite pairs shown in Figure 6. This suggests that random instrument errors and a difference in ozone caused by a difference in observation time between the satellites are relatively small. Standard deviations greater than 3–4% can be seen only at high latitudes. Larger values are likely related to a date error near the International Date Line. The date for the gridded GOME data was assigned using the date information from the start of the orbit. This may lead to a shift of 1 day for part of the orbit when the date change occurs somewhere in the middle of the orbit. This error should be kept in mind

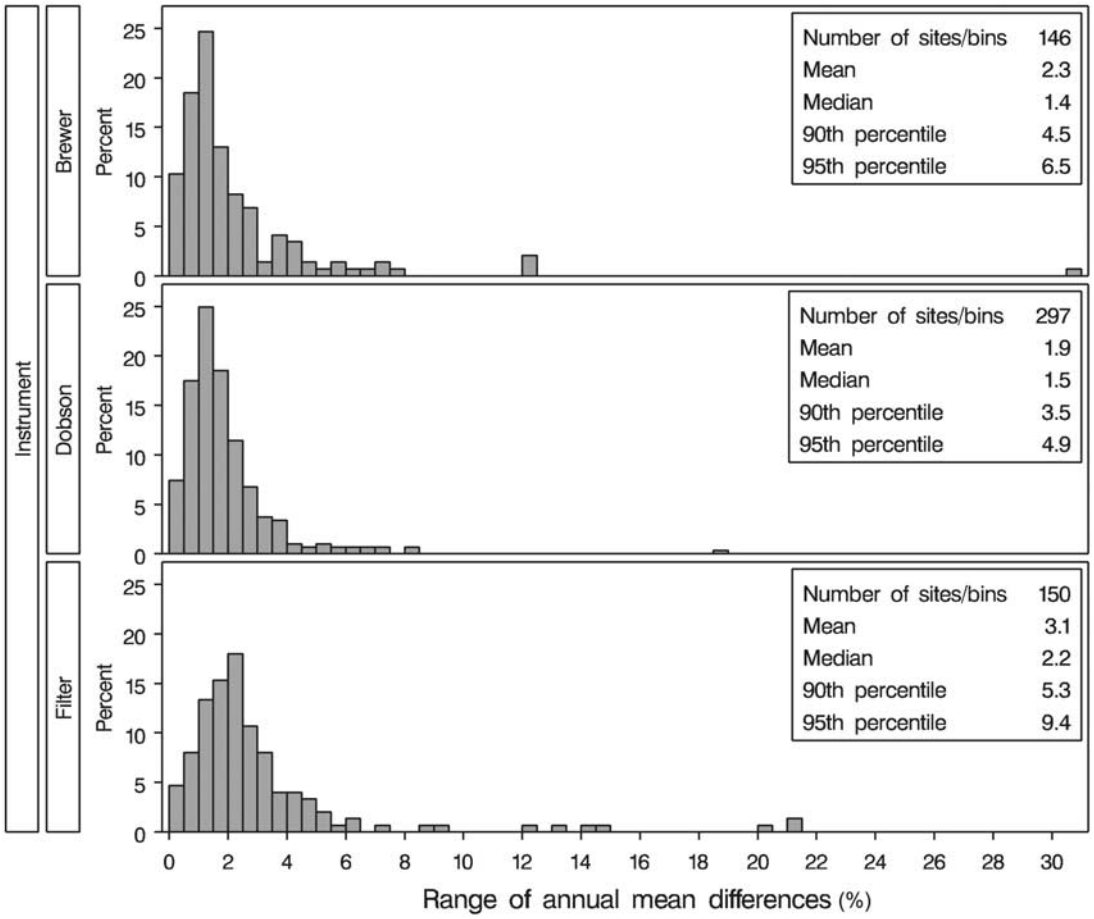

Figure 8. Histogram of distribution for the range of annual mean differences between ground-based DS observations and matched satellite overpasses for Dobson, Brewer, and filter instrument sites. The range of annual mean differences was estimated for each site and for each 5-year bin separately and then estimates from all bins were lumped together to produce the histogram.

when satellite data are compared with ground-based measurements at a few sites located near the 180th meridian.

3.3. Comparison Between Different Satellite Instruments Limited to Overpass Events

[34] Since our goal is to estimate the overall performance of the ground-based network by using satellite data as a "reference", it is important to know the accuracy of the "reference". To estimate its accuracy we extracted Meteor 3 TOMS, GOME, and OMI satellite overpass data that matched available Dobson, Brewer, and filter instrument DS and ZS observations and then compared these extracted overpasses with data from two satellite instruments with the longest records (Nimbus 7 and EP TOMS). The same five characteristics of the difference as used for the ground-based data (see section 2) were calculated for the satellite overpasses. The comparison was done for 3 bins: from 1991 to 1993 for the Nimbus 7 and Meteor 3 TOMS comparison, from 1996 to 2000 for the EP TOMS-GOME comparison and for 2001–2005 for the EP TOMS comparison with GOME and OMI. For each bin, the five characteristics were calculated for each site and then the site values from all bins were lumped together. Only stations located between 60°S and 60°N were used to make the results comparable to those for ground-based versus satellite differences discussed later in section 3.4. Sites located near the International Date Line where differences in satellite observation dates were detected (see section 3.2) were excluded from the daily standard deviation of the difference statistics. The comparison was done separately for overpasses corresponding to clear-sky or broken-cloud conditions (matching the days of DS observations on the ground) and overcast clouds (ZS observations on the ground).

[35] The results are summarized in Table 2. The mean difference between overpasses from different satellites for clear-sky or broken-cloud conditions is typically inside the ±1.5% range. Daily standard deviations are typically about 2%, and they rarely exceed 2.6%; this is in line with the results shown in Figure 6. Uncertainties related to longer-term variations of the difference (and expressed by the standard deviation of monthly differences, the seasonal amplitude of the difference, and the range of annual values) are on the order of 2%. The differences are larger for overcast conditions than for clear-sky or broken-cloud conditions. This reflects larger uncertainty in satellite

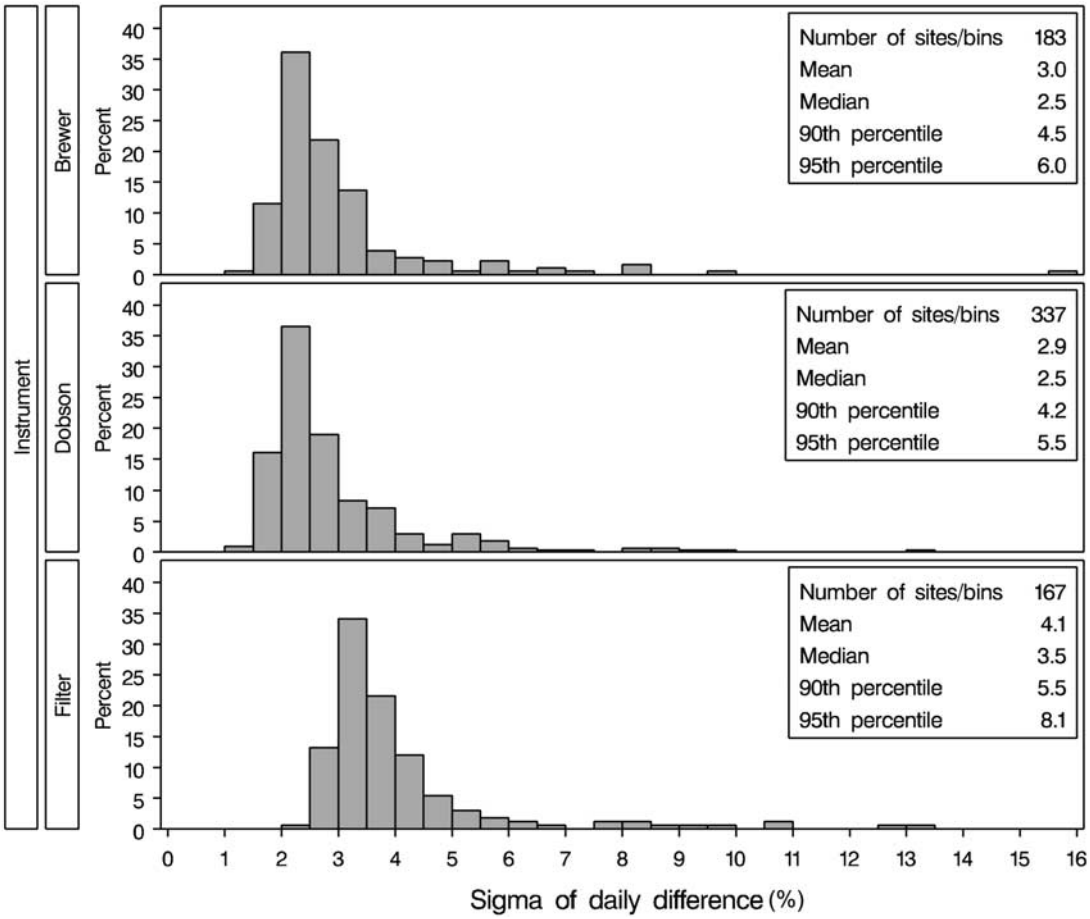

Figure 9. Histogram of distribution for the standard deviation of daily differences between ground-based DS observations and matched satellite overpasses for Dobson, Brewer, and filter instrument data. The standard deviation was estimated for each site and for each 5-year bin separately and then estimates from all bins were lumped together to produce the histogram.

retrievals under overcast conditions most likely due to a rapidly changing atmosphere under these conditions.

[36] There are obviously some differences between individual bins as indicated in Figure 5. There is also a difference in the results from overpasses for Dobson, Brewer and filter instrument sites that reflects the difference in station location, but these differences, while statistically significant, are relatively small (e.g., within ±0.6% for the mean difference). Therefore Table 2 presents a good overview of satellite observation uncertainties.

3.4. Comparison of Satellite Overpasses With Ground-Based Data

[37] The distribution characteristics of the difference between ground-based and satellite data for both DS and ZS measurements are summarized in Table 3, while Figures 7–12 show the histograms of the five characteristic distributions for DS observations only. We also limit this study to 60°S–60°N to avoid potential errors in satellite and ground-based data in the polar regions. If satellite overpasses from more than one satellite were available, then their average value was used. It is important to note that the study of the distribution of ground-based site performance (all characteristics) was done for five bins and then the results were lumped together unless otherwise specified.

3.4.1. Mean Difference

[38] Figure 7 shows the histogram of the ground-minus-satellite difference for Dobson, Brewer, and filter instrument data. The mean and median values of the difference for Brewer, Dobson and filter instrument stations are small, within ±0.7%. However, there are sites with a much larger mean differences. There are several potential explanations for a larger difference. The bias between ground-based and satellite data can be caused by station elevation. While satellite algorithms account for altitude, they do it for the average altitude of a large satellite ground pixel, which could be quite different from the actual altitude of a ground-based station. Mauna Loa is an extreme example of a high-altitude site that has a difference with satellite overpasses due to this reason. The Dobson instrument there is located at 3500 m above sea level and Dobson measurements are 4.4% lower than the gridded satellite data for that area. Mauna Loa and another high-altitude station, Kodaikanal, India, were excluded from the mean difference statistics.

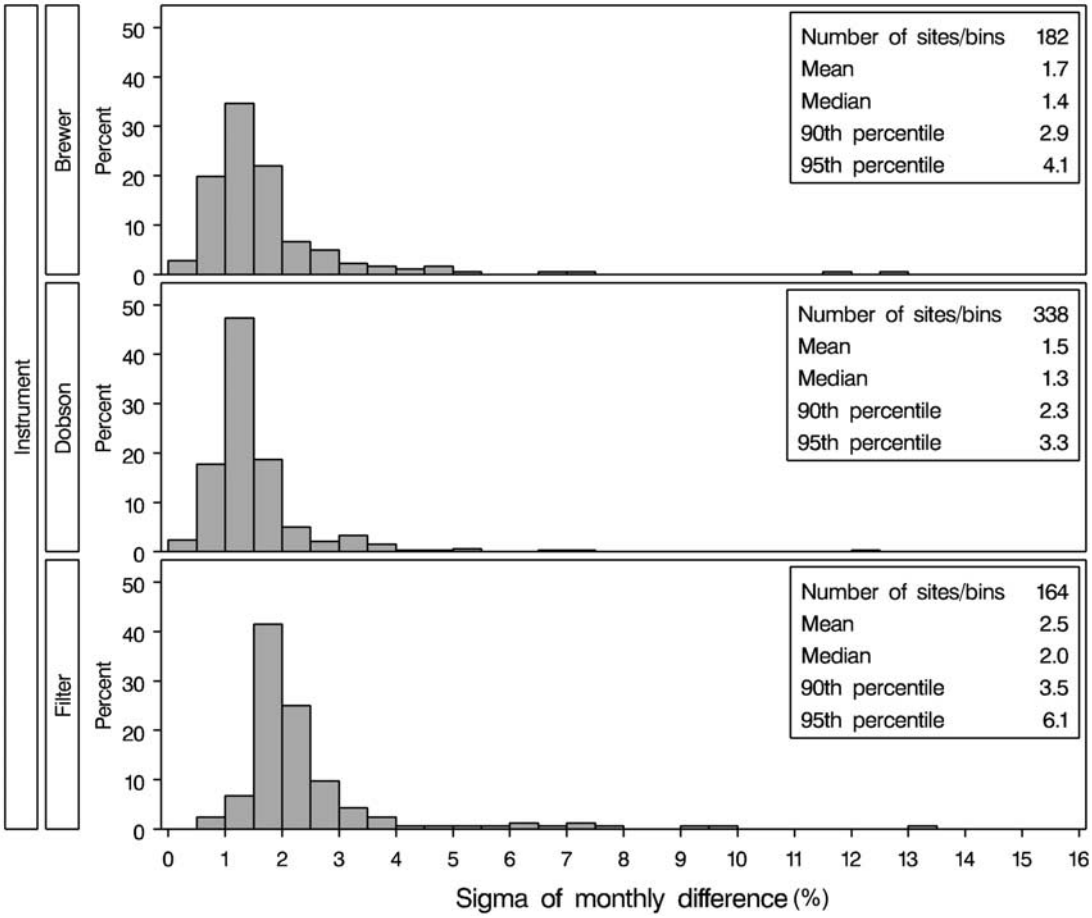

Figure 10. Histogram of distribution for the standard deviation of monthly mean differences between ground-based DS observations and matched satellite overpasses for Dobson, Brewer, and filter instrument data. The standard deviation was estimated for each site and for each 5-year bin separately and then estimates from all bins were lumped together to produce the histogram.

[39] The local condition at the site can also contribute to the difference. *Kerr and McElroy* [1995], demonstrated up to a 2% difference between early morning and afternoon total ozone values at a site located near a large urban center. This difference is related to ozone in the lower troposphere. Therefore daily mean values could be biased against satellite overpass data collected in the afternoon. This illustrates that some systematic differences between ground-based and satellite data can be caused by local conditions. However, as Figure 7 shows, biases outside the ±3% range are rare and may be indications of potential instrumental problems.

[40] The median value of the mean difference is within ±0.7% for both DS and ZS Dobson observations. Approximately one half of all Brewer sites have their ZS observation biased by 2.5% or more. This was previously seen in Figure 2. Brewer ZS measurements could overestimate ozone under heavy cloud conditions due to multiple scattering and ozone absorption of the UV radiation within the clouds. Some of these measurements could get through Brewer data quality control tests and affect daily averages. However, a 2.5% bias seems to be too large to be explained by this. The Brewer ZS algorithm is based on an empirical relationship between zenith radiation and column ozone. Most likely, the problem is related to the fact that many Brewer sites use the default coefficients for this relationship included in the Brewer operational software to calculate ZS ozone, rather then calculating an instrument-specific and site-specific set of coefficients after installation and any major refurbishment of the instrument [*Muthama et al.*, 1995]. Further analysis of individual Brewer ZS observations (instead of just daily mean values on days without DS observations that are presently available from the WOUDC) is required to clarify the origin of this bias.

3.4.2. Range of Annual Mean Differences

[41] While some systematic differences between ground-based and satellite data can be caused by local conditions, it is unlikely that these conditions are changing substantially with time. For this reason, the range of annual mean differences, i.e., the largest difference minus the smallest one, was used as another criterion for ground-based data validation. As Figure 8 shows, the median values of the range for Brewer and Dobson sites are nearly identical, 1.4% and 1.5% respectively, if only DS data are used. The 90th and 95th percentiles are larger for the Brewer instruments reflecting a relatively large number of "outliers".

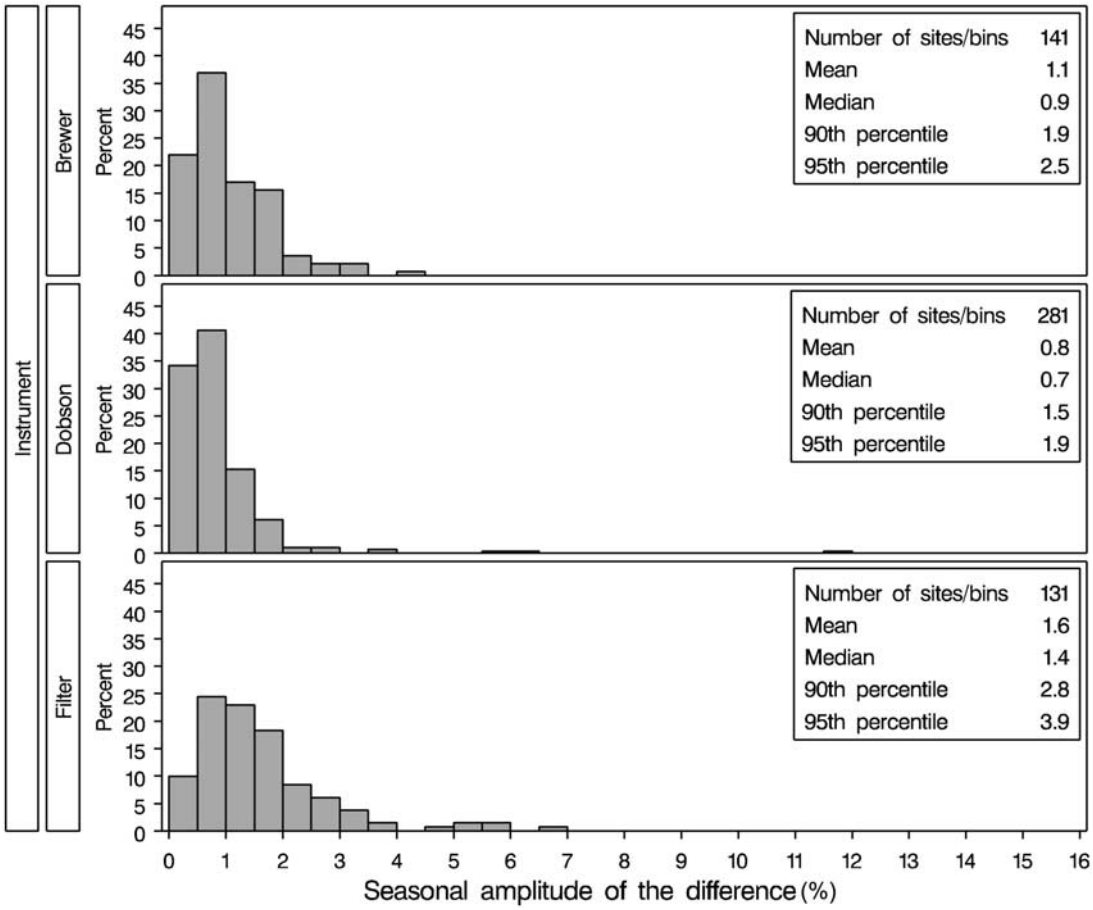

Figure 11. Histogram of distribution for the amplitude of the annual cycle of the difference between ground-based DS observations and matched satellite overpasses for Dobson, Brewer, and filter instrument data. The amplitude was estimated for each site and for each 5-year bin separately and then estimates from all bins were lumped together to produce the histogram.

Because of these outliers, the mean value of the range for Brewer instruments is also larger than that of the Dobson instruments (2.3% and 1.9% respectively). In one extreme case, the range for one of the Brewer sites was as large as 30% suggesting major problems with the instrument. Estimates for sites based on ZS data only show, in general, similar characteristics to those for DS conditions.

3.4.3. Standard Deviation of Daily Differences

[42] The precision of daily values is reflected by the standard deviation of daily differences. There are many sources of the difference between ground-based and satellite data and a large difference is not necessarily related to an error in either measurement. The distribution of standard deviations is nearly identical for Dobson and Brewer sites (DS data only) with the median value of 2.5% and mean of 3% (Figure 9). The 90th percentile values are also nearly identical (~4.5%), while the 95th percentile is larger for Brewer instruments than for Dobson instruments. The histogram of the difference between filter ozonometer data and satellite measurements is clearly different from these for Dobson and Brewers and the mean and median values are larger (4.1% and 3.5% respectively). As expected, the standard deviations calculated using ZS data only are noticeably higher, more than 4% for all types of instruments [see also *Fioletov et al.*, 1999].

3.4.4. Standard Deviation of Monthly Mean Differences

[43] This characteristic is sensitive to both short-term and long-term differences between ground-based and satellite data. Dobson and Brewer instruments demonstrated a similar performance for DS measurements with the median value of 1.3% and mean of 1.6% (Figure 10). Similar to the daily standard deviations and other statistics of the difference, filter ozonometers demonstrated larger mean and median values of the difference with satellite observations (2.5% and 2.0% respectively). Comparison of data from individual bins revealed that the main characteristics of the Dobson-satellite difference distribution remained the same for all 5 bins with the mean and median values virtually unchanged over the entire time period. The mean and median of monthly difference between ground-based and satellite observations for the Brewer instrument were small in the first bin (1979–1985). There were only 7 Brewer instruments reporting data at that time, and their data were carefully verified and reevaluated. The mean and median values are larger for the Brewer instrument than for the

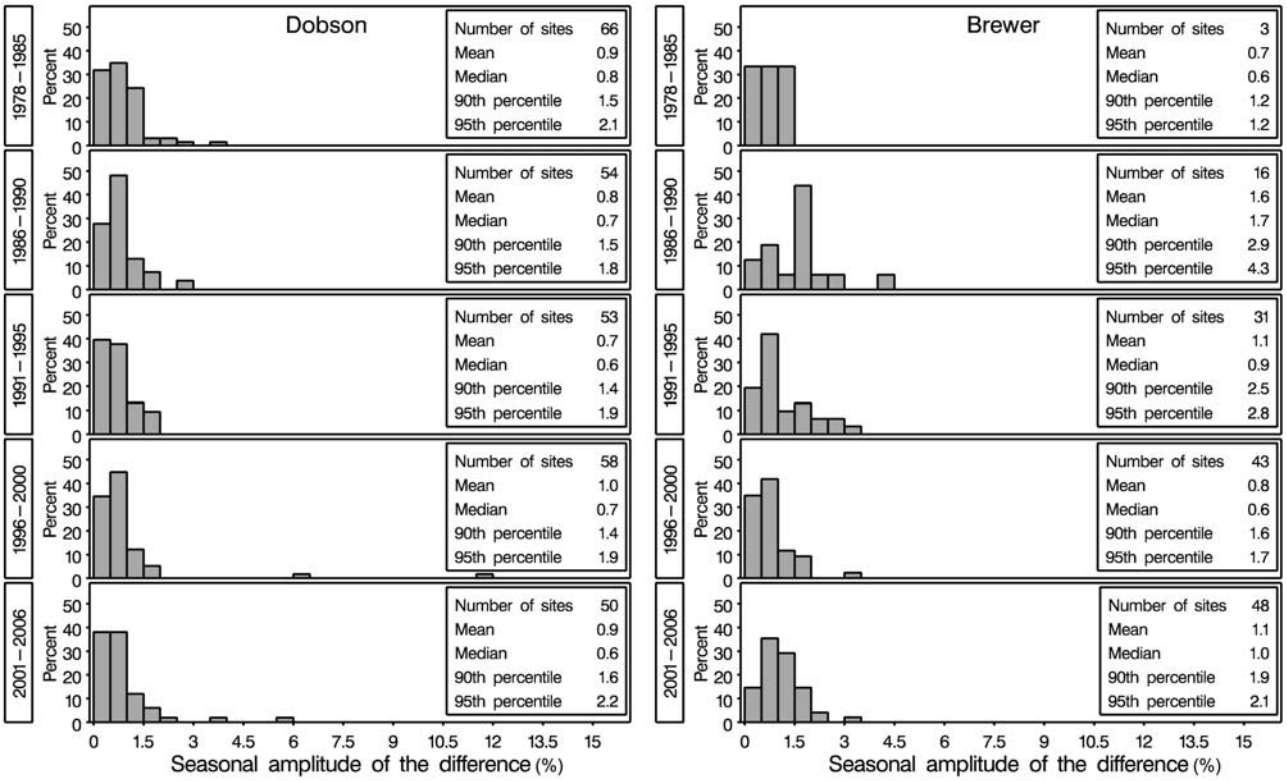

Figure 12. Histogram of distribution for the amplitude of the annual cycle of the difference between ground-based DS observations and matched satellite overpasses for Dobson, Brewer, and filter ozonometer data for different bins. The amplitude of the annual cycle of the Dobson-satellite difference is nearly the same for all 5 bins. The amplitude of the annual cycle of the Brewer-satellite difference was larger in 1980–1995 than in 2000–2006.

Dobson instrument during the next two bins (1986–1995). During the last bin, the mean and median values for Brewer sites are nearly the same as for Dobson sites. The mean and median values for filter ozonometer (not shown here) do not exhibit substantial differences between the five bins, except for the 1996–2000 period. The mean and standard deviations as well as the 90th percentile in that bin were the largest. The standard deviations of monthly mean differences calculated for ZS measurements are greater then those for DS data only.

3.4.5. Seasonal Amplitude of the Difference

[44] A seasonal cycle in the difference between ground-based and satellite observations can be caused by a difference in ozone retrieval algorithms. The present Dobson and Brewer instrument algorithms do not account for changes in the ozone absorption coefficients due to the annual cycle in ozone-weighted mean temperature caused by seasonal changes in the vertical distribution of ozone and temperature. Typically, these effects are small, except for extreme cases in the Antarctic winter and spring [*Bernhard et al.*, 2005]. Instrumental stray light [e.g., *Olafson and Asbridge*, 1981] as well as scattered light entering the instrument field of view during a DS measurement [*Josefsson*, 1992] can also yield a season-dependent difference in the difference between ground-based and satellite data. As mentioned in section 3.2, there are some seasonal differences in total ozone retrieved from different satellite observations. For the ground-based measurements, the seasonal component is likely caused by errors in the instrument extraterrestrial constant. These constants are determined during the calibration of Dobson and Brewer instruments, but they can change with time. These changes can usually be traced (and corrected) based on regular standard lamp test results, but this is not always done at some stations.

[45] Figure 11 shows the histograms of the seasonal component of the difference for Dobson, Brewer and filter instrument sites and Figure 12 shows the histograms for Dobson and Brewer sites for individual bins. Median amplitudes of the differences between Dobson and Brewer and satellite observations appear to be smaller than those between individual satellite instruments (Table 2). This may suggest that the seasonal difference is not necessarily related to a problem with ground-based measurements. From Figure 11, it appears that the seasonal cycle for Brewer sites is larger than for Dobson sites. However, as Figure 12 shows, the characteristics of the distribution for Brewer sites were different for different bins, while they remain virtually unchanged for Dobson sites. The median seasonal amplitude of the Brewer-satellite difference was larger in 1986–1990 than in the following years. This may be related to the Brewer calibration procedure used at that time: the extraterrestrial constant and effective ozone absorption coefficients were transferred from the traveling standard. In recent years, the effective ozone absorption coefficients were estimated from an independent test and only the ETC was transferred from

Table 4. Thresholds for Statistical Characteristics of the Difference Between Ground-Based and Satellite Data[a]

Parameter	Suspect	Outlier
DS Observations		
1. Mean difference, %	±3	±4
2. Standard deviation of daily difference, %	4.5	6
3. Standard deviation of monthly difference, %	3	4
4. Seasonal amplitude of the difference, %	2	3
5. Range of annual mean differences, %	4	5
ZS Observations		
Mean difference, %	±4	±5
Standard deviation of the daily difference, %	6	7
Standard deviation of the monthly difference, %	4	5
Seasonal amplitude of the difference, %	2.6	3.2
Range of annual mean differences, %	4	5

[a]A site performance within a single bin is classified as a "suspect" or "outlier" if a particular characteristic is outside the corresponding limit.

the traveling standard, which made the ETC estimates more accurate.

[46] The median amplitude of the Brewer-satellite difference for the last bin (2001–2006) is slightly larger than for the previous one. As Figure 2 suggests, there is a strong annual cycle in the EP TOMS-Brewer difference. The EP TOMS instrument during this time period is known to have had a solar-zenith-angle-dependent error. This error has been primarily removed by intercalibrating with NOAA-14 SBUV (see http://toms.gsfc.nasa.gov/news/Corrected_EP_TOMS_README.pdf) but a small bias may remain. If EP TOMS data are excluded from the comparison, the median value drops from 1.0 to 0.5, however, the 90th and 95th percentile values remain the same. This suggests that the cases of large seasonal amplitude of the difference are most likely related to ground-based data rather than satellite ones.

3.5. Overall Performance of the Dobson, Brewer, and Filter Instrument Networks (DS Data Only)

[47] Estimates from Table 3, particularly the percentiles, can be used to identify potential problems at individual sites. Comparison of Tables 2 and 3 demonstrates that the percentile values for satellite- and ground-based differences are always larger than percentiles of characteristics of the differences between individual satellite instruments. In other words, large differences are most likely related to potential problems at ground-based sites. This can be used to evaluate site performance. We call site performance for a single bin, "suspect" or "outlier" if a certain characteristic of the ground-based to satellite difference is outside the limits shown in Table 4. These limits approximately correspond to 90th and 95th percentiles estimated for Dobson and Brewer sites. Then, for each site the number of "suspect" and "outlier" characteristics was calculated for all bins. All described characteristics were related to individual bins. An additional characteristic, the range of mean differences for all bins, was introduced to evaluate stability of a site over the entire period. If this range exceeds the values given in Table 4 for annual mean differences, the site was also marked as "suspect" or "outlier".

[48] The characteristics of the differences are not independent. For example, a large annual cycle of the difference likely yields a large standard deviation of the differences. Therefore it is expected that a single problem could be reflected by more than one characteristic. We identify a site as one with minor issues if its entire record has one to three "suspect" characteristics or one "outlier" and zero or one "suspect" characteristics. A greater number of "suspect" or "outlier" characteristics will mark a site as one with major issues. Although this approach is rather subjective, it seems appropriate since our goal is to compare stations to each other.

[49] The same criteria from Table 4 were applied to all sites and instrument types. Table 5 shows the number of stations with minor and major issues among all ground-based stations with long records (with data available from three or more bins). About 60% of Dobson stations located between 60°S and 60°N have no "suspect" or "outlier" characteristics. If the distribution of the site differences is purely random and characteristics of the difference are independent, there are only 7% of all sites that have a record with no "suspect" or "outlier" characteristics in all five bins. For sites with data in only three bins, this probability is about 20%.

[50] The percentage of Brewer stations with no "suspect" or "outlier" characteristics (46%) is slightly less than that for Dobson stations. About a quarter of all Brewer sites were characterized as "sites with minor issues". Typically it means that only a part of the entire record was affected by an error in instrument constants and proper reprocessing of data can improve data quality. The criteria listed in Table 4 were developed for Dobson and Brewer sites. Therefore it was expected that sites equipped with the less accurate filter instruments would not perform as well as the Dobson and Brewer sites under these criteria. Nevertheless, more then half of all filter instrument sites were classified as sites with no issues or minor issues. These sites can be potentially useful for satellite data validation, particularly over areas where no Dobson and Brewer sites are available.

[51] Figure 13 shows the distribution of "good" sites with no "suspect" or "outlier" characteristics in different bins. In each bin, more than 80% of all Dobson sites are in that category. This number is noticeably larger than the value of 61% in Table 5 which is the summary of the entire period. A single "suspect" event in one of five bins will mark the entire record as a "suspect" for Table 5. The percentage of "good" Dobson sites represent a small increase over time, except for the last bin, when for many sites, the latest intercomparison result has not been used for data reevaluation yet. The absolute number of Dobson stations has decreased as some stations stopped their observations in recent years.

Table 5. Total Number of Dobson, Brewer, and Filter Stations With Long Records (i.e., With Data From 3 or More Bins) and Number of Stations With Minor and Major Issues (See Text for Details)[a]

	Dobson	Brewer	Filter
Number of stations with data from 3 or more bins	69 (100%)	41 (100%)	39 (100%)
Within the range	42 (61%)	19 (46%)	11 (28%)
Minor issues	11 (16%)	10 (24%)	10 (26%)
Major issues	16 (23%)	12 (29%)	18 (46%)

[a]The percent of the total number of stations with long records is given in parentheses.

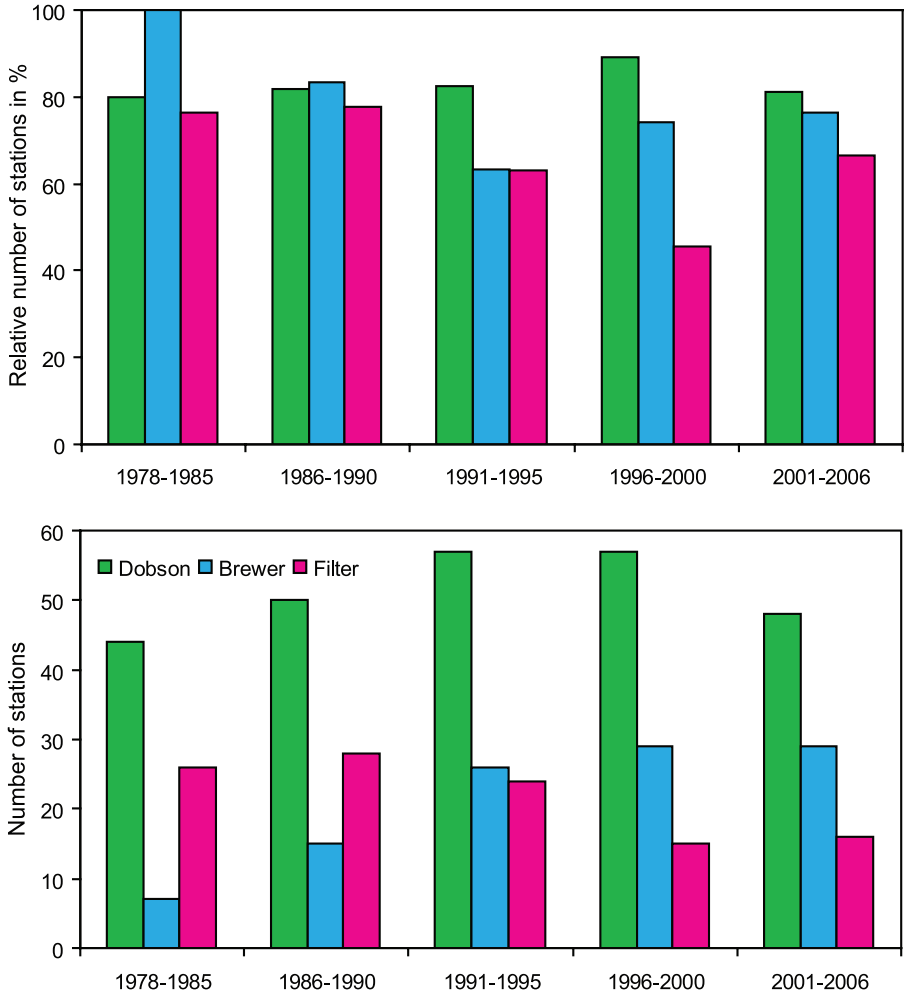

Figure 13. Absolute and relative (in percent from the total number) number of sites with "no issues" (see text) in the record in 5 bins for Dobson, Brewer, and filter instrument sites located between 60°S and 60°N.

[52] All seven Brewer sites demonstrated no "suspect" or "outlier" characteristics in the first bin. However, the rate of "good" stations declined in the third bin when the majority of Brewer sites started their records. Since then, the percentage of "good" stations has increased and it was 76% in the last bin. Filter ozonometer sites performed quite well in the first two bins due to rigid quality control applied by the staff at the MGO: records from unreliable instrument/sites were not submitted to the WOUDC. The number of "good" filter ozonometer sites has declined substantially in the fourth bin. It appears the percentage of "good" sites increased in the last bin, but the majority of filter sites have not submitted their recent data yet.

[53] As mentioned in section 3.4, high-latitude stations were excluded from the summary results shown in Table 4. It is expected that larger discrepancies will be seen between ground-based and satellite data at these latitudes and such discrepancies are not necessarily related to problems with the ground-based data. None of the Antarctic stations, with data available from at least three bins, can be considered as "good" based on the formal criteria from Table 4. With one exception, Antarctic stations have daily standard deviations and seasonal amplitude of the difference above the "suspect" limits at least in one of the five bins. Two Dobson stations in the Antarctic fall under the "Minor issues" category and four Dobson and one Brewer station are under the "Major issues" category. In the Arctic, three Dobson and two filter ozonometer sites are "within the range". Three sites of each instrument type fall under the "Minor issues" category and two sites of each instrument type fall under the "Major issues" category in the 60°N–90°N belt.

[54] It should be emphasized that the classification of stations as "good", "suspect", or "outlier" describes only the past station performance based on formal criteria described above and should not be automatically applied to the future data. Stations showing problems in the past can become "good" stations after reevaluation of the data, for example, if more calibration information becomes available and is applied. A discrepancy between ground-based and satellite data that have low probability according to the Table 4 criteria does not necessarily suggest a problem with the ground-based instrument. It could be related to some local problem with satellite data, as discussed in section 3.2. Or, keeping in mind a large number of ground-based stations and long periods of their operation, it could be even a random event. The suggested

classification is based on a statistical approach and it should be used only as a diagnostic tool to identify potential problems with the data. Correction of these problems should be based on internal procedures keeping the satellite and ground-based observing systems independent.

4. Summary and Discussion

[55] Comparison of total ozone data from different satellite instruments analyzed in this study demonstrates that the differences are typically within ±2% and very rarely are outside the ±3% envelope. The difference depends upon the season and region. The latter means that a change in satellite instruments or algorithms could be misinterpreted as a local "trend" if a record merged from multiple instrument data is analyzed. The magnitude of this artificial "trend" is not negligible since the overall observed total ozone decline over northern midlatitude is only about 3% [*WMO*, 2007].

[56] A good agreement between ground-based and satellite data on a small set of high quality sites does not guarantee that satellite data have the same good performance everywhere, even at the same latitudes. Particularly large, up to 3%, longitudinal differences between various satellite instruments can be seen in the tropical region. Very few ground-based stations are located in the tropics which makes it difficult to validate satellite observations for these latitudes.

[57] Dobson sites demonstrated a steady performance when their DS measurements were compared to satellite observations. The characteristics of the difference with satellites remained virtually unchanged over all 5 bins analyzed in this study. More then 60% of the Dobson sites do not have characteristics of the difference with satellite observations that are classified as "outliers" in this study.

[58] Lack of quality control is probably the main issue with some Brewer sites. While Brewer and Dobson sites demonstrated similar mean and median characteristics of the differences with satellite data (for DS data), the number of Brewer sites with large differences with satellite data is greater than those for Dobson sites. This is also reflected by the number of sites with minor and major issues in the record (Table 5 and Figure 13). It should be mentioned that recalibration and operation procedures of the Brewer network are less standardized compared to the Dobson network [*Staehelin et al.*, 2003]. The situation was particularly bad in the early 1990s, when Brewer observations started at a large number of sites. However, in most cases, Brewer data can be reprocessed using updated calibration constants to provide high-quality data.

[59] Differences between different satellite instruments, and between ground-based and satellite instruments under cloudy conditions are larger than for clear-sky conditions. The differences are particularly large for Brewer ZS measurements: the majority of Brewer sites have a bias of 2.4% or more against satellite instruments. This bias is likely related to the use of the default set of coefficients used for the ZS calculations at some sites. Brewer ZS observations at these sites should be reprocessed.

[60] Comparison with satellite data demonstrated again that filter instruments have lower precision compared to Dobson and Brewer instruments. The standard deviations of daily and monthly differences and other characteristics of the difference with satellite data are typically higher than those for Dobson and Brewer instruments. However, more than 76% of all filter ozonometer sites demonstrated no significant problems when compared with satellite overpasses during the 1980s, i.e., demonstrated performance that was just slightly worse than that for Dobson and Brewer sites at that time. This number dropped to 45% in 1996–2000. It appears the filter ozonometer network performance is slightly better in 2001–2006, however, a relatively small fraction of filter ozonometer data from that time interval have been submitted to the WOUDC thus far.

[61] In nearly 90% of all cases, Dobson and Brewer sites demonstrated 5-year mean differences with satellites to be within ±3% (for DS observations). Since instrument-related errors are typically independent at individual sites, the mean or median value of several sites can be used as a highly accurate reference for satellite measurements. For example, the seasonal median difference between all Brewer DS measurements at 25°–60°N and GOME and OMI overpasses remained within ±0.5% over a period of more than 10 years. It is essential to maintain ground-based measurements with high data quality through regular instrument intercomparisons and calibrations. A better understanding of sources of the difference between different types of ground-based and satellite instruments with possible adjustment of ozone retrieval algorithms is also required for data validation near 1% level. In particular, inadequate knowledge of the cross-section spectrum on ozone and its temperature dependence may be in part responsible for some of the absolute and seasonally changing discrepancies.

[62] Ground-based stations where total ozone measurements have deviations of more than ±3% against satellite observations may not be useful for the analysis of long-term changes in stratospheric ozone or comparisons with numerical simulations since available satellite instruments can measure ozone better than that. One obvious recommendation emerging from this study is to put more effort into improving the precision of ozone measurements at such stations in order to make them more suitable for ozone monitoring purposes.

References

Balis, D., et al. (2007a), Ten years of GOME/ERS2 total ozone data—The new GOME data processor (GDP) version 4: 2. Ground-based validation and comparisons with TOMS V7/V8, *J. Geophys. Res.*, *112*, D07307, doi:10.1029/2005JD006376.

Balis, D., M. Kroon, M. E. Koukouli, E. J. Brinksma, G. Labow, J. P. Veefkind, and R. D. McPeters (2007b), Validation of Ozone Monitoring Instrument total ozone column measurements using Brewer and Dobson spectrophotometer ground-based observations, *J. Geophys. Res.*, *112*, D24S46, doi:10.1029/2007JD008796.

Basher, R. E. (1982), Review of the Dobson spectrophotometer and its accuracy, *WMO Ozone Rep. 13*, World Meteorol. Organ., Geneva.

Basher, R. E. (1994), Survey of WMO-sponsored Dobson spectrophotometer intercomparisons, *WMO Ozone Rep. 19*, World Meteorol. Organ., Geneva.

Bernhard, G., R. D. Evans, G. J. Labow, and S. J. Oltmans (2005), Bias in Dobson total ozone measurements at high latitudes due to approximations in calculations of ozone absorption coefficients and air mass, *J. Geophys. Res.*, *110*, D10305, doi:10.1029/2004JD005559.

Bojkov, R. D., C. Mateer, and A. Hanson (1988), Comparison of ground-based and total ozone mapping spectrometer measurements used in assessing the performance of the global ozone observing system, *J. Geophys. Res.*, *93*, 9525–9533.

Bojkov, R. D., V. E. Fioletov, and A. M. Shalamjansky (1994), Total ozone changes over Eurasia since 1973 based on reevaluated filter ozonometer data, *J. Geophys. Res.*, *99*, 22,985–22,999.

Bramstedt, K., J. Gleason, D. Loyola, W. Thomas, A. Bracher, M. Weber, and J. P. Burrows (2003), Comparison of total ozone from the satellite instruments GOME and TOMS with measurements from the Dobson network 1996–2000, *Atmos. Chem. Phys.*, *3*, 1409–1419.

Coldewey-Egbers, M., M. Weber, L. N. Lamsal, R. de Beek, M. Buchwitz, and J. P. Burrows (2005), Total ozone retrieval from GOME UV spectral data using the weighting function DOAS approach, *Atmos. Chem. Phys.*, *5*, 5015–5025.

De Muer, D., and H. De Backer (1992), Revision of 20 years of Dobson total ozone data at Uccle (Belgium): Fictitious Dobson total ozone trends induced by sulfur dioxide trends, *J. Geophys. Res.*, *97*, 5921–5937.

Dobson, G. M. B. (1957), Dobson Ozone Spectrophotometer - Observers' Handbook for the Ozone Spectrophotometer, *Ann. Int. Geophys. Year*, vols. V and XVI.

Eskes, H. J., R. J. van der A, E. J. Brinksma, J. P. Veefkind, J. F. de Haan, and P. J. M. Valks (2005), Retrieval and validation of ozone columns derived from measurements of SCIAMACHY on Envisat, *Atmos. Chem. Phys. Discuss.*, *5*, 4429–4475.

Fioletov, V. E., J. B. Kerr, E. W. Hare, G. J. Labow, and R. D. McPeters (1999), An assessment of the world ground-based total ozone network performance from the comparison with satellite data, *J. Geophys. Res.*, *104*, 1737–1747.

Fioletov, V. E., J. B. Kerr, C. T. McElroy, D. I. Wardle, V. Savastiouk, and T. S. Grajnar (2005), The Brewer reference triad, *Geophys. Res. Lett.*, *32*, L20805, doi:10.1029/2005GL024244.

Fleig, A. J., P. K. Bhartia, C. G. Wellemeyer, and D. S. Silberstein (1986), An assessment of the long-term drift in SBUV total ozone data, based on comparison with the Dobson network, *Geophys. Res. Lett.*, *13*, 1359–1362.

Gustin, G. P., S. A. Sokolenko, B. G. Dudko, and V. V. Lagutina (1985), Ozonometer M-124 (in Russian), *Tr. Glav. Geofiz. Obs. im. A. I. Voeikova*, *499*, 60–67.

Josefsson, W. A. P. (1992), Focused Sun observations using a Brewer ozone spectrophotometer, *J. Geophys. Res.*, *97*, 15,813–15,817.

Kerr, J. B. (2002), New methodology for deriving total ozone and other atmospheric variables from Brewer spectrophotometer direct sun spectra, *J. Geophys. Res.*, *107*(D23), 4731, doi:10.1029/2001JD001227.

Kerr, J. B., and C. T. McElroy (1995), Total ozone measurements made with the Brewer ozone spectrophotometer during STOIC 1989, *J. Geophys. Res.*, *100*, 9225–9230.

Kerr, J. B., I. A. Asbridge, and W. F. J. Evans (1988), Intercomparison of total ozone measured by the Brewer and Dobson spectrophotometers at Toronto, *J. Geophys. Res.*, *93*, 11,129–11,140.

Kerr, J. B., C. T. McElroy, and D. W. Wardle (1997), The Brewer instrument calibration center 1984–1996, in *Atmospheric Ozone: Proceedings of the XVIII Ozone Symposium in L'Aquila*, edited by R. D. Bojkov and G. Visconti, pp. 915–918, Parco Sci. and Tecnologico D'Abruzzo, Italy.

Komhyr, W. D. (1980), Operations handbook: Ozone observations with a Dobson spectrophotometer, in *WMO Ozone Rep. 6*, World Meteorol. Organ., Geneva.

Komhyr, W. D., R. D. Grass, and R. K. Leonard (1989), Dobson spectrophotometer 83: A standard for total ozone measurements 1963–1987, *J. Geophys. Res.*, *94*, 9847–9861.

McPeters, R. D., and G. J. Labow (1996), An assessment of the accuracy of 14.5 years of Nimbus 7 TOMS Version 7 ozone data by comparison with the Dobson Network, *Geophys. Res. Lett.*, *23*, 3695–3698.

McPeters, R. D., J. R. Herman, P. K. Bhartia, A. J. Krueger, C. G. Wellemeyer, and G. J. Labow (1998), Earth Probe Total Ozone Mapping Spectrometer (TOMS) data products user's guide, *NASA Reference Publication 1998-206895*, National Aeronautics and Space Administration, Goddard Space Flight Center, Greenbelt, Maryland.

Muthama, N. J., U. Scimia, A. M. Siani, and S. Palmieri (1995), Toward optimizing Brewer zenith sky total ozone measurements at the Italian stations of Rome and Ispra, *J. Geophys. Res.*, *100*, 3017–3022.

Olafson, R. A., and I. A. Asbridge (1981), Stray light in the Dobson spectrophotometer and its effect on ozone measurements, in *Proc. Quadrenn. Int. Ozone Symp.*, edited by J. London, pp. 46–47, IAMAP, Boulder, Colo.

Staehelin, J., A. Renaud, J. Bader, R. McPeters, P. Viatte, B. Högger, V. Bugnion, M. Giroud, and H. Schill (1998), Total ozone series of Arosa (Switzerland). Homogenization and data comparison, *J. Geophys. Res.*, *103*, 5827–5841.

Staehelin, J., J. Kerr, R. Evans, and K. Vanicek (2003), Comparison of total ozone measurements of Dobson and Brewer spectrophotometers and recommended transfer functions, *GAW (Global Atmosphere Watch) Rep. 149*, World Meteorol. Organ. (WMO), Geneva, 2003.

Stolarski, R. S., and S. Frith (2006), Search for evidence of trend slowdown in the long-term TOMS/SBUV total ozone data record: The importance of instrument drift uncertainty and fingerprint detection, *Atmos. Chem. Phys.*, *6*, 4057–4065.

Van Roozendael, M., et al. (2006), Ten years of GOME/ERS2 total ozone data–The new GOME Data Processor (GDP) Version 4: 1. Algorithm description, *J. Geophys. Res.*, *111*, D14311, doi:10.1029/2005JD006375.

Vanicek, K. (2006), Differences between ground Dobson, Brewer and satellite TOMS-8, GOME-WFDOAS total ozone observations at Hradec Kralove, Czech, *Atmos. Chem. Phys.*, *6*, 5163–5171.

Weber, M., L. N. Lamsal, M. Coldewey-Egbers, K. Bramstedt, and J. P. Burrows (2005), Pole-to-pole validation of GOME WFDOAS total ozone with ground-based data, *Atmos. Chem. Phys.*, *5*, 1341–1355.

World Meteorological Organization (WMO) (1993), Handbook for ozone data reevaluation, *WMO Ozone Rep. 29*, Geneva.

World Meteorological Organization (WMO) (2007), Scientific assessment of ozone depletion: 2006, *WMO Ozone Rep. 50*, Geneva.

R. Evans, NOAA Earth System Research Laboratory/Global Monitoring Division, 325 Broadway, Boulder, CO 80305-3328, USA.

V. E. Fioletov, E. W. Hare, C. T. McElroy, and V. Savastiouk, Environment Canada, ARQX, 4905 Dufferin Street, Toronto, ON, Canada M3H 5T4. (vitali.fioletov@ec.gc.ca)

U. Köhler, Meteorological Observatory Hohenpeissenberg, German Weather Service, Albin-Schwaiger-Weg 10, D-82383 Hohenpeissenberg Hohenpeissenberg, Germany.

G. Labow, SSAI-Lanham and NASA Goddard Space Flight Center, 10210 Greenbelt Road, Suite 400, Lanham, MD 20706, USA.

K. Miyagawa, Ozone and Radiation Division, Aerological Observatory, Japan Meteorological Agency, 1-2, Nagamine, Tsukuba 305-0052, Japan.

A. Redondas, Izana Observatory, National Institute of Meteorology, E-38071 Santa Cruz de Tenerife, Spain.

A. M. Shalamyansky, Main Geophysical Observatory, 7 Karbyshev Str., 194018, St. Petersburg, Russia.

J. Staehelin, Institute for Atmospheric and Climate Science, ETH Zurich, 8092 Zürich, Switzerland.

K. Vanicek, Solar and Ozone Observatory, Czech Hydrometeorological Institute, Hvezdarna 456, 500 08 Hradec Kralove 8, Czech Republic.

M. Weber, Institute of Environmental Physics and Remote Sensing, University of Bremen, Otto-Hahn-Allee 1, D-28359, Bremen, Germany.