JOURNAL OF GEOPHYSICAL RESEARCH, VOL. 111, D14305, doi:10.1029/2005JD006718, 2006

Intercomparison of GOME, ozonesonde, and SAGE II measurements of ozone: Demonstration of the need to homogenize available ozonesonde data sets

Xiong Liu,[1] Kelly Chance,[1] Christopher E. Sioris,[1] Thomas P. Kurosu,[1] and Michael J. Newchurch[2]

Received 27 September 2005; revised 6 March 2006; accepted 24 March 2006; published 21 July 2006.

[1] We investigate the large systematic biases, especially in the stratosphere, between ozone profiles retrieved from the Global Ozone Monitoring Experiment (GOME) and ozonesonde observations at some ozonesonde stations. GOME retrievals are intercompared with both ozonesonde data at 33 stations between 75°N and 71°S and Stratospheric Aerosol and Gas Experiment II (SAGE II) data during 1996–1999. GOME stratospheric column ozone (SCO) over the altitude range ~15–35 km usually agrees with SAGE II SCO to within 2.5 DU (1.5%, 1 DU = 2.69×10^{16} molecules cm^{-2}) without significant spatiotemporal dependence but is systematically larger than ozonesonde SCO by 8–20 DU (5–10%) over carbon iodine (i.e., an ozonesonde technique) stations and most stations within 30°N–30°S. Evaluation of GOME, SAGE II, TOMS, and Dobson data here demonstrates that those biases mainly originate from ozonesonde underestimates in the stratosphere. GOME retrievals also show large positive biases of 20–70% at carbon iodine stations (except for Syowa) and most stations within 30°N–30°S over ~10–20 km, where ozone concentration is low, while the biases relative to SAGE II data over ~15–20 km is usually 10–20%. The discrepancies over this altitude region reflect biases in GOME retrievals as well as ozonesonde measurements. In addition, GOME/sonde biases in both SCO and profiles (especially in the lower stratosphere and upper troposphere) vary from station to station and depend on sonde technique, instrument type, sensor solution, and data processing, demonstrating the need to homogenize available ozonesonde data sets and standardize future operational procedures for reliable and consistent satellite validation.

Citation: Liu, X., K. Chance, C. E. Sioris, T. P. Kurosu, and M. J. Newchurch (2006), Intercomparison of GOME, ozonesonde, and SAGE II measurements of ozone: Demonstration of the need to homogenize available ozonesonde data sets, *J. Geophys. Res.*, *111*, D14305, doi:10.1029/2005JD006718.

1. Introduction

[2] Ozone profiles can be measured through in situ or remote-sensing devices. Because no single instrument measures ozone at all altitude ranges with an adequate spatiotemporal coverage, observations from various instruments must be combined to obtain the four-dimensional (4-D) (3-D space plus time) distribution of ozone [*World Meteorological Organization* (*WMO*), 1998]. In practice, observations from different instruments are often intercompared with each other. For instance, ozonesonde observations are usually used to validate satellite observations.

[3] We have developed an algorithm to retrieve ozone profiles from the Global Ozone Monitoring Experiment (GOME) [*Liu et al.*, 2005]. The retrievals show large systematic biases especially in the stratosphere relative to ozonesonde observations at some stations and these biases vary from station to station, as detailed in sections 3.2 and 3.3. To investigate the sources of these biases and evaluate GOME retrievals, we intercompare GOME retrievals with both ozonesonde measurements within the latitude range 75°N–71°S and Stratospheric Aerosol and Gas Experiment II (SAGE II) data from 1996 through 1999. GOME provides ozone profiles globally (~2500 profiles/day, global coverage in 3 days at the equator) but with coarse vertical resolutions of 7–12 km at 20–40 km and of 9–16 km in the troposphere [*Liu et al.*, 2005]. The ozonesonde is the primary tool for monitoring ozone in the troposphere and lower stratosphere. It measures profiles at high vertical resolution (~100–200 m) from the surface up to 35 km, but with limited and uneven spatiotemporal coverage. SAGE measures ozone profiles at ~1 km vertical resolution but with limited geographical coverage (~30 profiles/day) due to the solar occultation technique used. SAGE data have been widely used for deriving stratospheric ozone trends

[1] Atomic and Molecular Physics Division, Harvard-Smithsonian Center for Astrophysics, Cambridge, Massachusetts, USA.
[2] Atmospheric Science Department, University of Alabama in Huntsville, Huntsville, Alabama, USA.

0148-0227/06/2005JD006718$09.00

[*Randel et al.*, 1999; *Cunnold et al.*, 2000; *Newchurch et al.*, 2000, 2003] because of their long data record (1979 to present) and high accuracy (especially with recent updates of the retrieval algorithms). Because of the constant viewing geometry, the SAGE ozone retrieval quality hardly depends on the measurement location unless affected by volcanic aerosols and clouds. Thus SAGE data can be used to assess the spatial consistency between both GOME and ozonesonde data. However, because of many fewer coincidences [*Wang et al.*, 2002] between SAGE II and ozonesonde than those between GOME and ozonesonde, it is difficult to make a direct intercomparison of ozonesonde and SAGE II data at a particular ozonesonde station. The comparison of GOME with both ozonesonde and SAGE II data serves as an indirect intercomparison.

[4] Since we have already validated GOME Tropospheric Column Ozone (TCO) against ozonesonde TCO and compared GOME ozone profiles with SAGE II data above ~15 km [*Liu et al.*, 2005], here we compare only GOME stratospheric column ozone (SCO) with ozonesonde and SAGE II data and compare GOME ozone profiles with ozonesonde data. Because ozonesonde only samples up to ~35 km and the measurement accuracy is usually reduced above ~25 km, ozonesonde SCO does not actually cover the whole stratosphere and is less accurate. A more accurate SCO quantity can be determined by subtracting the TCO from the concurrently measured total ozone [*Wozniak et al.*, 2005]. However, to show the heterogeneity in ozonesonde data, we choose to validate GOME and ozonesonde SCO within the same altitude range (~15–35 km). We introduce the three data sets and describe the comparison methodology in section 2. Section 3 compares GOME with SAGE II with respect to SCO (section 3.1), and versus ozonesonde in terms of SCO (section 3.2) and profiles (section 3.3). We summarize our results in section 4.

2. Data and Methodology

2.1. Global Ozone Monitoring Experiment (GOME) and GOME Ozone Profile Retrievals

[5] GOME, launched in April 1995 on board the European Space Agency's second Earth Remote Sensing satellite, measures radiances backscattered from the Earth's atmosphere in the wavelength region 240–790 nm. The local equator crossing time is 10:30 am in the descending mode. Because of its moderate spectral resolution (0.2–0.4 nm) and high signal-to-noise ratio in the ozone absorption bands, the vertical distribution of ozone down through the troposphere can be retrieved [*Munro et al.*, 1998; *Hoogen et al.*, 1999; *Hasekamp and Landgraf*, 2001; *van der A et al.*, 2002; *Müller et al.*, 2003; *Liu et al.*, 2005].

[6] Our algorithm to retrieve ozone profiles from GOME ultraviolet spectra has been described in detail in a previous paper [*Liu et al.*, 2005]. The retrieved profiles have 11 layers, with the National Centers for Environmental Prediction/National Center for Atmospheric Research (NCEP/NCAR) reanalysis tropopause defining the boundary between two adjacent layers in the stratosphere and troposphere; each layer is ~5 km (4–6 km) thick except for the top layer (~10 km). The troposphere is divided into two or three equal log-pressure layers depending on the location of the tropopause. In addition to ozone profiles, the output contains total, stratospheric, and tropospheric column ozone (i.e., TO, SCO, and TCO) and their error estimates. The accuracy (i.e., random, smoothing, and other errors) of the retrievals is estimated to be 20–30% in the troposphere and lower stratosphere and 5–10% in the middle and upper stratosphere. The accuracy of TO, SCO, and TCO, are estimated to be 1.6%, 2.3%, 21% on average, respectively. The retrieved TO agrees with observations from the Total Ozone Mapping Spectrometer (TOMS) and Dobson/Brewer instruments to within 6 DU (2%, 1 DU = 2.69×10^{16} molecules cm^{-2}) at most of the locations. TCO agrees well with ozonesonde TCO, with mean biases (MBs) mostly <3 DU (15%) and 1σ standard deviations (SDs) <3–8 DU (13–27%) [*Liu et al.*, 2005]; the global distribution of TCO agrees well with the GEOS-CHEM simulation over most regions of the globe [*Liu et al.*, 2006]. We also compared our retrieved ozone profiles above ~15 km against SAGE II data during 1996–1999; the MBs and SDs are usually within 15% [*Liu et al.*, 2005].

[7] Since the previous study [*Liu et al.*, 2005], we have found and corrected the precision estimate (i.e., the previous overestimate of the measurement error by a factor of ~2.3) in the GOME level 1 data (289–307 nm) after June 1998 due to the channel 1a/1b boundary change from 307 nm to 282 nm, which on average leads to an underestimate in TCO of 1–2 DU, an overestimate in SCO of 0–3 DU, but an overestimate of the retrieved ozone at two layers over ~10–20 km in the tropics by ~20–40% (~1–3 DU at each layer). However, we have only applied this correction to the subset data coincident with ozonesonde observations, and have not yet applied it to global retrievals due to the small effect on SCO, and the time-consuming global retrievals.

2.2. Ozonesondes

[8] We use ozonesonde data from 33 stations (see Figure 1 and Table 1) during 1996–1999. Data was primarily obtained from the World Ozone and Ultraviolet Data Center (WOUDC, http://www.woudc.org). Data unavailable or incomplete at WOUDC are directly obtained from the data originators (http://croc.gsfc.nasa.gov/shadoz; http://www.cmdl.noaa.gov) [*Oltmans et al.*, 2001; *Thompson et al.*, 2003] as shown in Table 1. The measurements were made with three types of ozonesonde: the electrochemical concentration cell (ECC), Brewer-Mast (BM), and carbon iodine (CI). Some ECC stations changed instruments (e.g., EnSci versus SPC), sensing solutions (e.g., 1% KI buffered versus 2% KI unbuffered), or preparation procedures [*Johnson et al.*, 2002; *Thompson et al.*, 2003]. For example, four National Oceanic and Atmospheric Administration/Climate Monitoring and Diagnostics Laboratory (NOAA/CMDL) stations (Boulder, Hilo, American Samoa, and Tahiti) changed the sensor solution from 1% buffered (1% b) to 2% unbuffered (2% ub). It is a general practice to scale measured ozone profiles to independently measured TO [*Logan*, 1999, and references therein]. However, an altitude-independent correction may distort the shape of the vertical profiles and lead to errors in the troposphere [*Hilsenrath et al.*, 1986], so not all the ozonesonde stations adopt this procedure. As shown in Table 1, all the BM and CI stations (except for Java) adopt this scaling procedure; most of the ECC stations within 30°N–30°S do not apply it but most of the other ECC stations do.

ECC sondes to be 3–5%, better than those of BM and CI sondes (5–15%) [*Smit and Kley*, 1998]. The SHADOZ measurement precision is also estimated to be ±5% [*Thompson et al.*, 2003]. Systematic biases exist in the CI sondes relative to ECC sondes, but the CI sondes are usually self-consistent (J. A. Logan, personal communication, 2004). The ozonesonde performance significantly depends on model types, sensor solutions, Pump Correction Factors (PCFs), background signal removal, and individual preparation procedures [*Smit and Kley*, 1998; *WMO*, 1998; *Johnson et al.*, 2002; *Thompson et al.*, 2003; *Smit and Sträter*, 2004a, 2004b]. It was found in both JOISE-1998 and JOISE-2000 that although both SPC-6A and EnSci-Z overestimate tropospheric ozone by 5–10% below 20 km, they show significant differences above 20 km; SPC-6A shows a mean bias from +5% at 25 km to −8% at 35 km while EnSci-Z shows a ∼+10% offset over 25–35 km [*Smit and Sträter*, 2004a, 2004b]. The JOISE-2000 discloses that ECC ozonesondes with 1% b gives ∼5% and ∼10% higher ozone that those with 0.5% half buffered and 2% ub, respectively [*Smit and Sträter*, 2004b]. The field dual-flight tests performed by NOAA/CMDL showed that ozone concentration with the use of 1% b is higher by 5–15% than that with 2% ub [*Johnson et al.*, 2002].

2.3. Stratospheric Aerosol and Gas Experiment II (SAGE II)

[10] SAGE II uses the solar occultation technique to measure the attenuated solar radiation through the Earth's limb in seven channels from 0.385 to 1.02 μm during each sunrise and sunset [*Mauldin et al.*, 1985]. We use version 6.2 SAGE II ozone profiles down to ∼15 km (http://www-sage2.larc.nasa.gov). It was shown that the measurements of version 5.9 have an accuracy of ∼5% over 20–45 km [*Cunnold et al.*, 1989, 1996; *Wang et al.*, 1996]. The version 6.1 algorithm mainly improves the retrievals below 20 km; the average agreement between version 6.1 SAGE II and ozonesondes is ∼10% down to 15 km [*Wang et al.*, 2002], with coincidence criteria of ±2° in latitude, ±12° in longitude, and ±24 hours in time. Version 6.2 data are similar to version 6.1 data with differences between them usually less than 0.5% (http://www-sage2.larc.nasa.gov).

2.4. Comparison Methodology

[11] We first compare GOME and SAGE II SCO at seven 20°-latitude bins from 70°N to 70°S and around each ozonesonde station. The criteria for comparison with SAGE II data are: within the same day, ±1.5° latitude, and ±600 km in longitude. SAGE II profiles are integrated into subcolumns according to GOME retrieval grids for the top eight GOME retrieval layers (∼15–60 km) and convolved with GOME retrieval averaging kernels to the GOME vertical resolution since SAGE II's vertical resolution is much higher than that of GOME. Because SAGE II data below ∼15 km are not used, we do not use the averaging kernels below layer 4 in the convolution, which essentially uses climatological a priori ozone profiles to complement SAGE II data below this altitude. The SAGE II SCO ($SCO_{4\text{-}7}$ within layers 4–7 or ∼15–35 km and $SCO_{8\text{-}11}$ within ∼35–60 km) is summed from the transformed SAGE II profiles. To compare GOME and SAGE II around each ozonesonde station, we use all coincidences (each coincidence still meets the above criteria) within ±5° latitude and ±40° longitude of each station.

[12] We then compare GOME SCO (column ozone from tropopause to the top-most GOME layer below ozonesonde burst altitude, ∼30 to 35 km), $SCO_{4\text{-}7}$, and profiles with ozonesonde data. We use only ozone profiles that extend above the sixth GOME retrieval layer (∼30 km). The coincidence criteria with ozonesonde measurements are the same as those with SAGE II data except at Kaashidhoo, Paramaribo, Kuala Lumpur, and Irene, where the longitude collocation criterion is relaxed to 12° in order to obtain enough collocations. We first integrate these profiles into six or seven subcolumns (∼0–30/35 km), complement the ozonesonde profiles with monthly mean SAGE II observations above ozone burst altitude, and then apply GOME retrieval averaging kernels and obtain the ozonesonde SCO and $SCO_{4\text{-}7}$.

[13] In the intercomparison, GOME data can be compared with both ozonesonde and SAGE II data over the ∼15–35 km altitude range (i.e., GOME retrieval layers 4 to 7). For all the comparisons, we remove outliers (1–2% of coincidences) which are defined as outside 3σ of the mean difference for each station or latitude range. To reduce the effects of clouds on GOME retrievals, we only use GOME data with cloud fraction less than 0.8. Using stricter cloud criteria will not affect the main results shown below. The relative differences are defined as: (GOME − ozonesonde or SAGE II)/(ozonesonde or SAGE II) × 100%.

3. Results and Discussion

3.1. Comparison of GOME and SAGE II Stratospheric Column Ozone

[14] Table 2 lists the GOME/SAGE II comparison statistics in $SCO_{8\text{-}11}$ and $SCO_{4\text{-}7}$ during 1996–1999 for seven 20°-latitude bins with a total of 16956 comparisons. Over the altitude range ∼35–60 km, there are negative MBs of 3–6 DU (8–16%) with poor correlation between these two data sets, because there exists a wavelength-dependent radiometric calibration error in the GOME normalized radiances between 289–307 nm and we do not use measurements below 289 nm, which provide information for ozone over this altitude range [*Liu et al.*, 2005]. For the ∼15–35 km altitude range, the MBs are within 2.5 DU (1.5%) with negligible latitude dependence and good correlation of 0.86–0.95. These biases are much smaller than those in individual layers due to opposite signs at different altitudes. The 1σ of the differences, varying from 4.5 DU

Table 2. GOME/SAGE II SCO Comparison Statistics for Seven Latitude Bins[a]

Latitude Range	Number of Comp.	Layers 8–11 (∼35–60 km)	Layers 4–7 (∼15–35 km)
70°N–50°N	4578	−3.7 ± 1.7, 0.82	−0.6 ± 10.9, 0.94
50°N–30°N	2953	−3.2 ± 1.9, 0.78	−2.3 ± 11.1, 0.92
30°N–10°N	828	−4.2 ± 2.0, 0.64	−0.1 ± 4.8, 0.94
10°N–10°S	259	−5.0 ± 2.6, 0.44	1.2 ± 4.5,0.90
10°S–30°S	997	−5.6 ± 2.8, 0.44	1.2 ± 6.2,0.86
30°S–50°S	3052	−5.3 ± 2.3, 0.56	−1.5 ± 9.5, 0.91
50°S–70°S	4289	−5.7 ± 2.6, 0.79	−1.1 ± 12.9, 0.90

[a]Number of comparisons, mean biases and 1σ standard deviations in DU, and correlation coefficients.

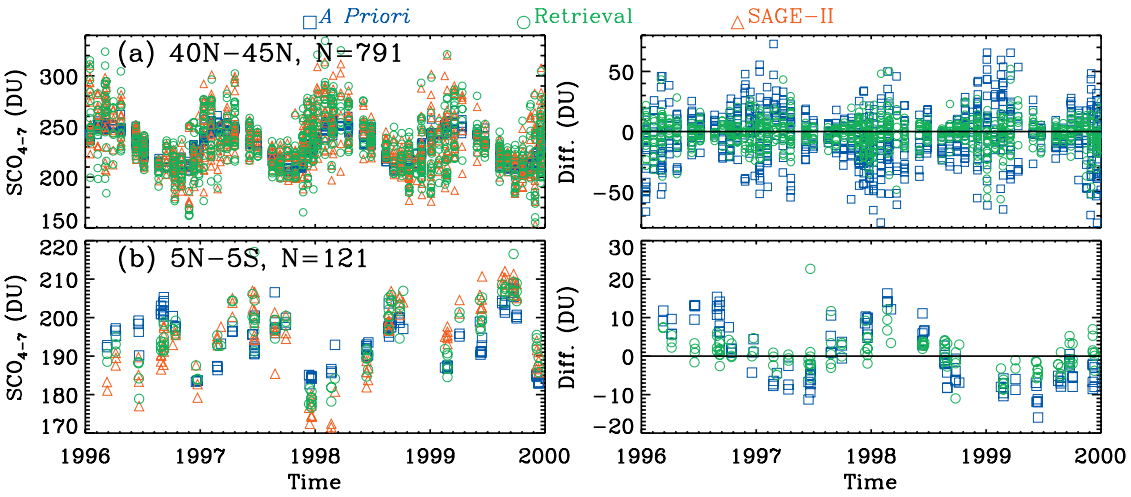

Figure 2. Comparison of Global Ozone Monitoring Experiment (GOME) retrieval/a priori and Stratospheric Aerosol and Gas Experiment II (SAGE II) column ozone over altitude range ~15–35 km (SCO_{4-7}) during 1996–1999. (left) Time series and (right) differences (GOME retrieval/a priori minus SAGE II).

(2.4%) in the tropics to 12 DU (5.6%) at high latitudes, is due mainly to the latitudinal variation of the spatiotemporal ozone variability [*Allen and Reck*, 1997] and the sampling biases between GOME and SAGE II (i.e., nadir versus limb). Figure 2 shows two examples of time series of comparisons at 40°–45°N and 5°N–5°S. GOME SCO_{4-7} agrees well with SAGE data; their differences are much less scattered than those between a priori and SAGE SCO_{4-7} and do not show significant temporal drift. In 40°–45°N, some large biases occur during the winter and spring when ozone variabilities are strong. Differences of >10 DU in 5°N–5°S are caused by the South Atlantic Anomaly (SAA), which can introduce large retrieval errors [*Liu et al.*, 2005]. Figure 3 shows the MBs and SDs between GOME and SAGE II SCO_{4-7} around each ozonesonde station, with 19 to 612 comparisons at each station (Table 1). The correlation coefficients are greater than 0.85 (Table 1) except for Ascension Island due to the SAA. The MBs are within ±2.5 DU (1.5%) except for negative biases of 6–7 DU (<3%) over three Northern European stations. Because of the closeness of these three stations, some coincidences are included at all the locations, leading to similar biases. Our previous evaluation of GOME retrievals indicates that GOME TCO is 5–8 DU larger than GEOS-CHEM and MOZAIC TCO during November–February at Frankfurt and it is anticorrelated with GEOS-CHEM TCO over Northern Europe [*Liu et al.*, 2006]. The biases over this region may result from the large surface albedo variability and the difficulty in discriminating snow/ice and clouds.

3.2. Comparison of GOME and Ozonesonde Stratospheric Column Ozone

[15] Figure 3 also shows MBs and SDs of GOME/sonde differences in SCO and SCO_{4-7} with the number of comparisons (from 6 to 285) and correlation coefficients shown in Table 1. Figure 4 shows examples of comparisons among a priori, retrieved and ozonesonde SCO (left) at selected stations (arranged by latitudes) and their differences (right). At Hohenpeißenberg and Lauder (Figures 4a and 4e), the

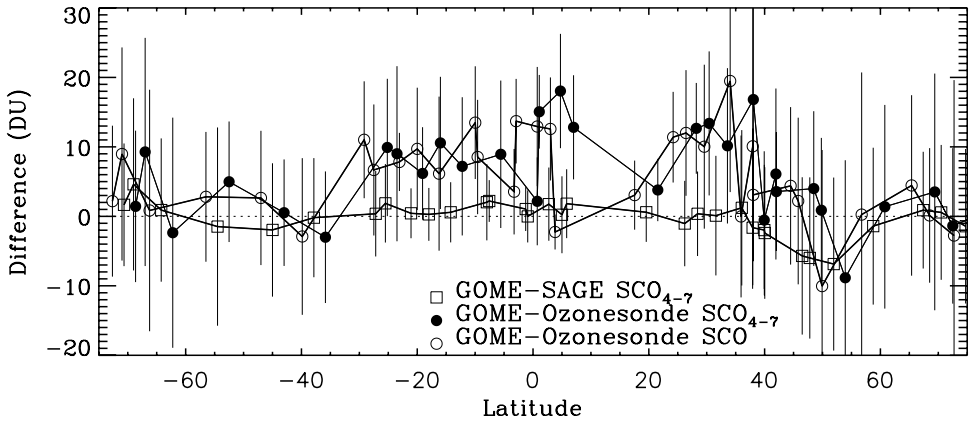

Figure 3. Mean biases and 1σ standard deviations for comparing GOME and SAGE II SCO_{4-7} (column ozone from ~15 to 35 km), GOME and sonde SCO_{4-7}, and GOME and sonde SCO (from tropopause to ~30/35 km) at ozonesonde stations during 1996–1999. The *x* axes for solid and open circles are shifted by +2° and −2°, respectively, for clarity.

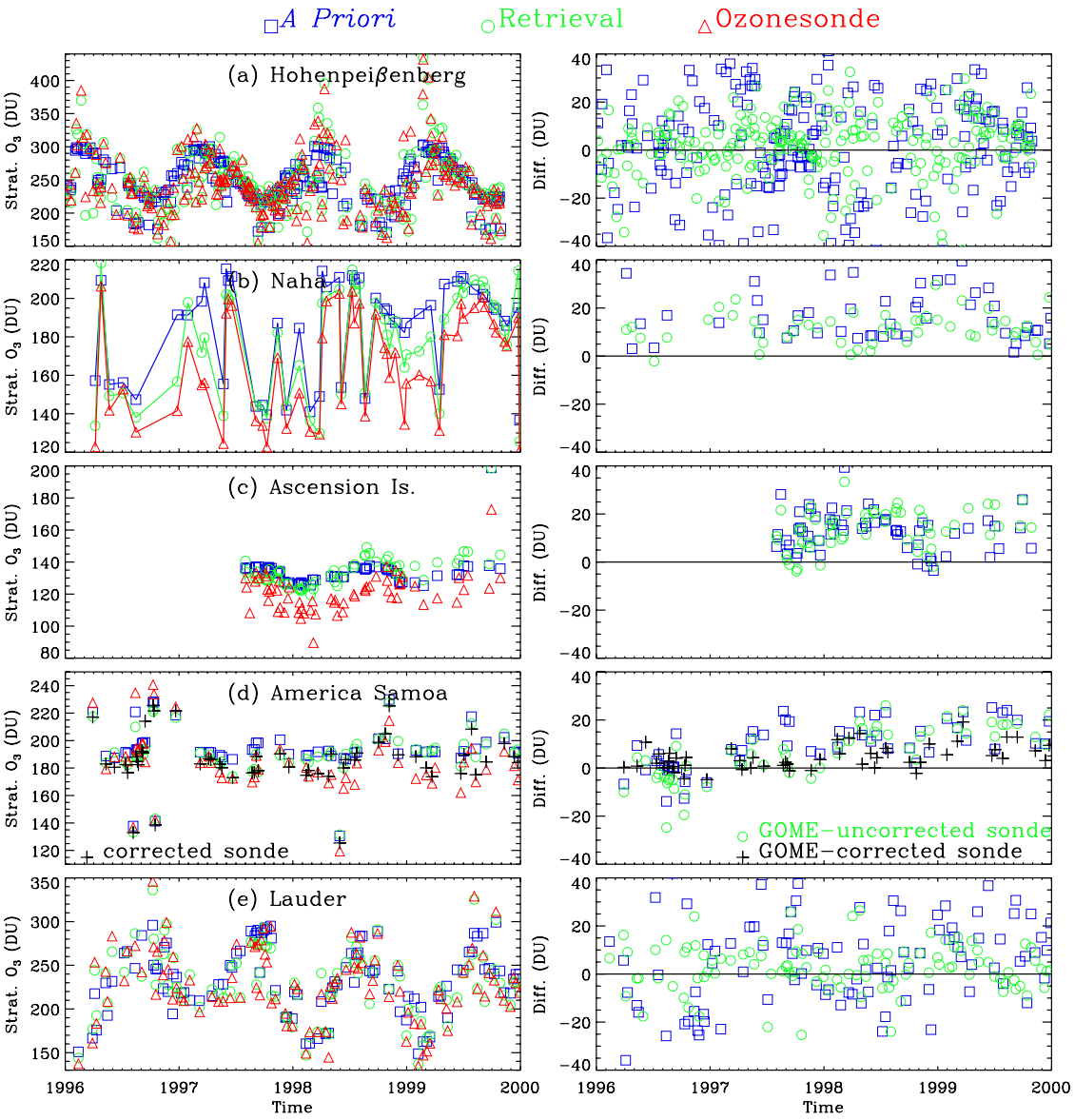

Figure 4. Comparisons of GOME/a priori and ozonesonde stratospheric column ozone (SCO, column ozone between tropopause and the top-most layer below ozonesonde burst) at (left) five selected stations and (right) the differences (GOME/a priori−ozonesonde). The plus symbols in Figure 4d show similar comparison for ozonesonde data with altitude-dependent correction and total ozone normalization.

retrievals agree well with ozonesonde SCO and are significantly improved over the a priori values. The MBs are less than 4.5 DU (2%). As seen from Figure 3, similar good agreements are also found at most of the mid- and high-latitude (>30°N/S) stations (except for CI stations) and several low-latitude stations between 30°S and 30°N (e.g., Hilo, Paramaribo). However, positive biases of 9–19 DU (6–9%) exist over those CI stations (e.g., Figure 4b) and most of the low-latitude stations; obvious changes in the biases occur at several locations within the data period (Figures 4c and 4d). Unlike the MBs, which show large variations among stations, the SDs of the biases (for both SCO and $SCO_{4\text{-}7}$) typically range from 3–9 DU (2–6%) within 30°N/S to 9–20 DU (4–10%) at higher latitudes. Some large SDs or poor correlations in the tropics are due to the instrument/procedure change within the data period (e.g., American Samoa). The comparison statistics in $SCO_{4\text{-}7}$ are usually similar to those in SCO.

[16] GOME/sonde $SCO_{4\text{-}7}$ comparisons are usually similar to the GOME/SAGE II $SCO_{4\text{-}7}$ comparisons at middle to high latitudes (Figure 3). However, at Tateno and Kagoshima, GOME/sonde biases are larger by 10–15 DU; at Hohenpeißenberg and Payerne, GOME $SCO_{4\text{-}7}$ shows opposite biases with respect to ozonesonde and SAGE II $SCO_{4\text{-}7}$. In the tropics, the GOME/sonde biases vary from station to station. They are much more inhomogeneous than the GOME/SAGE II biases, with usually larger SDs, poorer correlation (Table 1), and large positive biases at most locations.

[17] Although ozonesonde measurements at the CI stations (except for Java) have been normalized to concurrent Dobson observations, GOME SCO values are still system-

Table 3. GOME/sonde Comparison Statistics[a]

Station	Switch Date	GOME-Ozonesonde SCO	
		Number	Mean Bias ± 1σ, R
Boulder	21 Aug 1997	92	3.1 ± 9.6, 0.96
		36	1.6 ± 6.5, 0.97
		55	4.5 ± 10.4, 0.96
Hilo	15 Apr 1998	53	3.1 ± 5.0, 0.98
		28	0.9 ± 4.6, 0.99
		25	5.4 ± 4.3, 0.99
American Samoa	17 Apr 1998	55	6.1 ± 11.1, 0.86
		32	−1.0 ± 8.4, 0.95
		23	16.0 ± 5.1, 0.96
American Samoa with normalization	17 Apr 1998	53	4.4 ± 5.0, 0.96
		31	2.6 ± 4.2, 0.98
		22	6.9 ± 5.1, 0.96
Tahiti	6 May 1998	48	9.7 ± 8.8,0.93
		24	4.1 ± 6.9,0.97
		24	15.3 ± 6.7,0.95

[a]Number of comparisons, mean biases, and 1σ standard deviations in DU and correlation coefficients in stratospheric column ozone (SCO) from tropopause to the top-most layer (layer 6 or 7, i.e., ∼30–35 km) below burst altitude for the whole period (first row), before (second row), and after (third row) switching sensor solutions from 1% KI buffered to 2% KI unbuffered.

atically higher. This contradicts the facts that GOME TO usually agrees well with Dobson TO and GOME TCO agrees well with ozonesonde TCO at those stations [*Liu et al.*, 2005]. This discrepancy may result from the large uncertainty in estimating ozone above burst during the normalization. *Thompson et al.* [2003] show that the estimated column ozone above burst with the constant mixing ratio extrapolation is ∼10–20 DU higher than that using the SBUV climatology for tropical stations. Thus overestimation of column ozone above burst could lead to a smaller or even opposite SCO correction and may explain some of the GOME/sonde SCO biases. Although a similar normalization procedure is used over the two BM stations (i.e., Hohenpeißenberg and Payerne), ozonesonde SCO agrees well with GOME SCO. This may be because the average mixing ratio measured over 10–8 hPa is extrapolated at these BM stations (R. Stübi, personal communication, 2005) while the last value before burst is extrapolated at CI stations, and is related to the altitude-dependent performance of these two techniques.

[18] Although no time-dependent drifts in the biases were found at most stations, a manifest transition in biases can be seen at stations with major instrumental changes. At Ascension Island, the mean GOME/sonde SCO bias increases by ∼7.5 DU since 1998 (Figure 4c), corresponding to the switch of sonde models from EnSci to mostly SPC in early 1998 (F. J. Schmidlin, personal communication, 2005). This bias change is consistent with previous findings that EnSCi measures 5–10% more ozone than SPC above 20 km [*Smit and Sträter*, 2004a, 2004b]. Figure 4d shows good agreement at American Samoa during 1996–1997 but GOME values are systematically and significantly larger. A transition occurs approximately in April 1998, when the sensor solution was changed from 1% b to 2% ub. Table 3 shows the biases for all, 1% b, and 2% ub measurements at four CMDL stations. At American Samoa and Tahiti, the biases for 2% ub data are larger by 17 and 11 DU than those for 1% b data, respectively. These bias changes are consistent with the findings of *Johnson et al.* [2002] and *Smit and Sträter* [2004b] that measured ozone with 1% b are ∼10% higher than that with 2% ub above 20 km. Although the sensor solutions were also switched at Boulder and Hilo, the bias changes are only 2.9 and 4.5 DU, respectively. This is because the data at these two stations are already homogenized by applying the altitude-dependent correction and scaling the extrapolated TO (with SBUV) to concurrent Dobson measurements of TO (S. J. Oltmans and B. J. Johnson, personal communication, 2004). Figure 4d (plus symbols) and Table 3 also show the SCO comparison at American Samoa with the above correction. This correction significantly improves the data homogeneity, reducing the 2% ub bias by 9 DU and increasing the 1% b bias by 3.6 DU. The bias change is reduced from 17 to 4 DU and the overall standard deviation is reduced from 11.1 to 5.0 DU. Sensor solutions or sonde techniques were also changed at La Réunion (from 1.0% b to 0.5% b in May 1998) and Java (from CI to ECC with 2% ub in August 1999). However, transitions in biases could not be determined due to few collocations for one of the periods.

[19] It was noted that the ozonesonde SCO with 1% b data agree better with GOME retrievals at the CMDL stations according to Figure 4 and Table 3 and the normalization with altitude-dependent correction mainly improves the 2% ub data. This seems to be inconsistent with the results of *Johnson et al.* [2002] who reported that 2% ub data agree better with accurate UV reference photometer data with the use of NOAA/CMDL PCFs than 1% b data, since both the 1% b and 2% ub data used are processed with the NOAA/CMDL measured PCFs. The better agreement with the 1% b measurements is consistent with the results of JOISE-2000 [*Smit and Sträter*, 2004b] conducted in an environmental simulation chamber. However, the JOISE-2000 conclusion is based on using the standard *Komhyr* [1986] PCFs. With the NOAA/CMDL PCFs, it would be expected that 2% ub measurements also performed better due to the large difference of 0.03–0.15 between NOAA/CMDL and standard Komhyr PCFs in the stratosphere [*Johnson et al.*, 2002].

[20] Eleven SHADOZ stations are included in our comparison. Except for Paramaribo (−2.2 DU) and Nairobi (3.6 DU) where there are good agreements, positive MBs between GOME and sonde SCO ranges from 6.7 DU (4.2%) at Irene to 15–16 DU (8–13%) at Ascension Island, Tahiti, and American Samoa during 1998–1999 (solid circles in Figure 5). Similar biases exist in the ozonesonde integrated column ozone (i.e., TCO+SCO, solid triangles in Figure 5) as the GOME/sonde biases in TCO are usually within 3.0 DU (open circles in Figure 5). The relatively large positive biases (5.2 DU) in TCO at Tahiti may be partly caused by ozonesonde measurements due to the use of 2% ub sensor solution (see discussion in next section). *Thompson et al.* [2003] found that the sonde-evaluated TO is higher than TOMS Version-7 TO by ∼5–10% (12–24 DU) at most of the SHADOZ stations except at Nairobi, Irene, and La Réunion, where the biases are relatively smaller [see *Thompson et al.*, 2003, Figure 9] (Paramaribo is not included). *Labow et al.* [2004] found that EP-TOMS Version-7 data are ∼2–3% higher than Dobson TO in the Southern Hemisphere. After removing the biases from the TOMS Version-7 data, the remaining biases of ∼6–18 DU

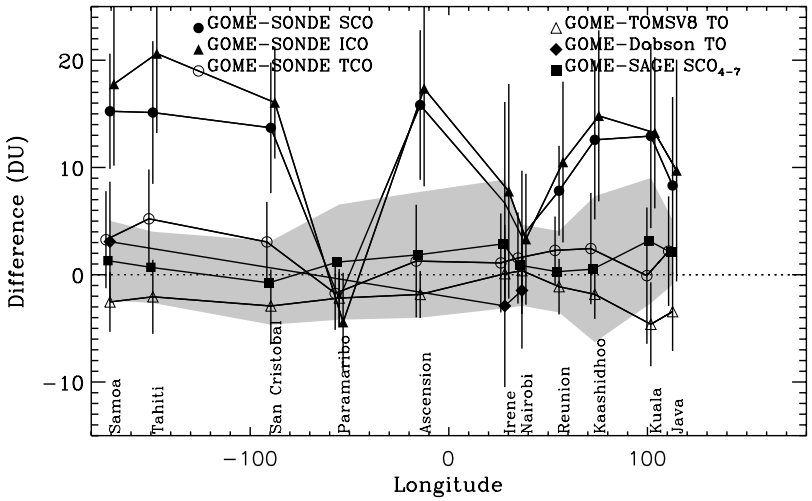

Figure 5. Mean biases and 1σ standard deviations between GOME and other measurements at 11 tropical stations during 1998–1999. SCO, stratospheric column ozone from tropopause to ∼30/35 km; ICO, sum of SCO and tropospheric column ozone (TCO); TO, total column ozone; $SCO_{4\text{-}7}$, SCO but from ∼15 to ∼35 km. The shaded area shows the ±1σ spread of GOME/SAGE II differences. Some *x* axes are slightly shifted for clarity.

are similar to the GOME/sonde SCO biases. Because the GOME measurements are usually consistent with TOMS Version-8, Dobson, SAGE II, ozonesonde TCO to within 3 DU (Figure 5), those large SCO biases of 10–16 DU, significant relative to 1σ of the various differences, mainly originate from ozonesonde measurements (i.e., stratospheric ozone is underestimated at those ozonesonde stations). Similarly, systematic ozonesonde underestimations in SCO also occur at CI stations, Santa Cruz, and Easter Island, where there are large GOME/sonde SCO biases.

3.3. Comparison of GOME and Ozonesonde Ozone Profiles

[21] Figure 6 shows the absolute and relative MBs and SDs between GOME and ozonesonde profiles for all the stations, arranged into five groups according to latitude (except for e). The absolute MBs generally show similar altitude-dependent patterns, usually larger at higher latitudes. The MBs are usually within 3 DU for the bottom two layers (∼0–10 km) and positive at layers 4 and 5 (∼15–25 km); they decrease and usually become negative from layers 6 or 7 on upward. There are large biases of 5–10 DU at layers 4 or 5 for most high-latitude stations (Figure 6a) and a few middle and low-latitude stations. This altitude-dependent bias pattern is similar to that between GOME and SAGE II, supporting that some biases arise from GOME retrievals. We believe that these systematic retrieval biases at individual layers are partly caused by the residual radiometric calibration error in the GOME normalized radiances below 307 nm after the second-order polynomial correction [*Liu et al.*, 2005]. Although large biases occur at individual layers for middle and high-latitude stations, the overall biases in SCO are small due to canceling errors (except for CI stations). Consistent positive biases over ∼15–35 km at some tropical stations lead to large positive SCO biases. The absolute SDs are mainly a function of latitude, with larger SDs at higher latitudes due to stronger ozone spatiotemporal variability. For example, at Syowa, which is located at the edge of the polar vortex (with maximum ozone outside the vortex and low ozone inside the vortex), the ozone in the ∼15–25 km layer can vary from 10 DU to more than 100 DU during the austral spring and there are sometimes large GOME/sonde differences during this period. Thus a small spatiotemporal mismatch may lead to large differences.

[22] The relative MBs and SDs are typically within 20% except for layers 3–4 (∼10–20 km) over three CI stations and most of the stations within 30°N/S. At CI stations (except for Syowa), the MBs and SDs are within 30–70%, usually larger than other stations at similar latitudes. This supports the notion that the CI measurements are usually noisier and systematic underestimates exist [*Smit and Kley*, 1998]. Except at Paramaribo and Easter Island, where the biases are small, stations within 30°N–30°S usually show large positive biases of 20–55% over ∼10–20 km, where ozone concentration is low (i.e., 5–10 DU over layer 3 and 8–15 DU over layer 4).

[23] Figure 7 shows the MBs and SDs in layers 4 and 5 (∼15–25 km) for comparing GOME and SAGE II/ozonesonde. The biases between GOME and SAGE II are averaged only from 1996 through May 1998 to avoid large retrieval errors in this altitude range in the global retrievals resulting from incorrect error estimate in GOME radiances between 282 and 307 nm after June 1998. At most middle and high-latitude non-CI stations, the MDs and SDs with respect to both ozonesonde and SAGE II are similar. Over the tropical and CI stations, the GOME/SAGE II MBs are usually homogenous, mostly less than 5% for layer 5 and within 8–20% for layer 4, but the GOME/sonde comparisons show larger positive biases (e.g., mostly 5–20% for layer 5 and 20–55% for layer 4), varying from station to station. The large GOME/ozonesonde bias of 33% for layer 5 at Syowa is related to the large ozone spatiotemporal variability and the large GOME/SAGE II spatial footprint. As seen from Figure 7, SAGE II values are also signifi-

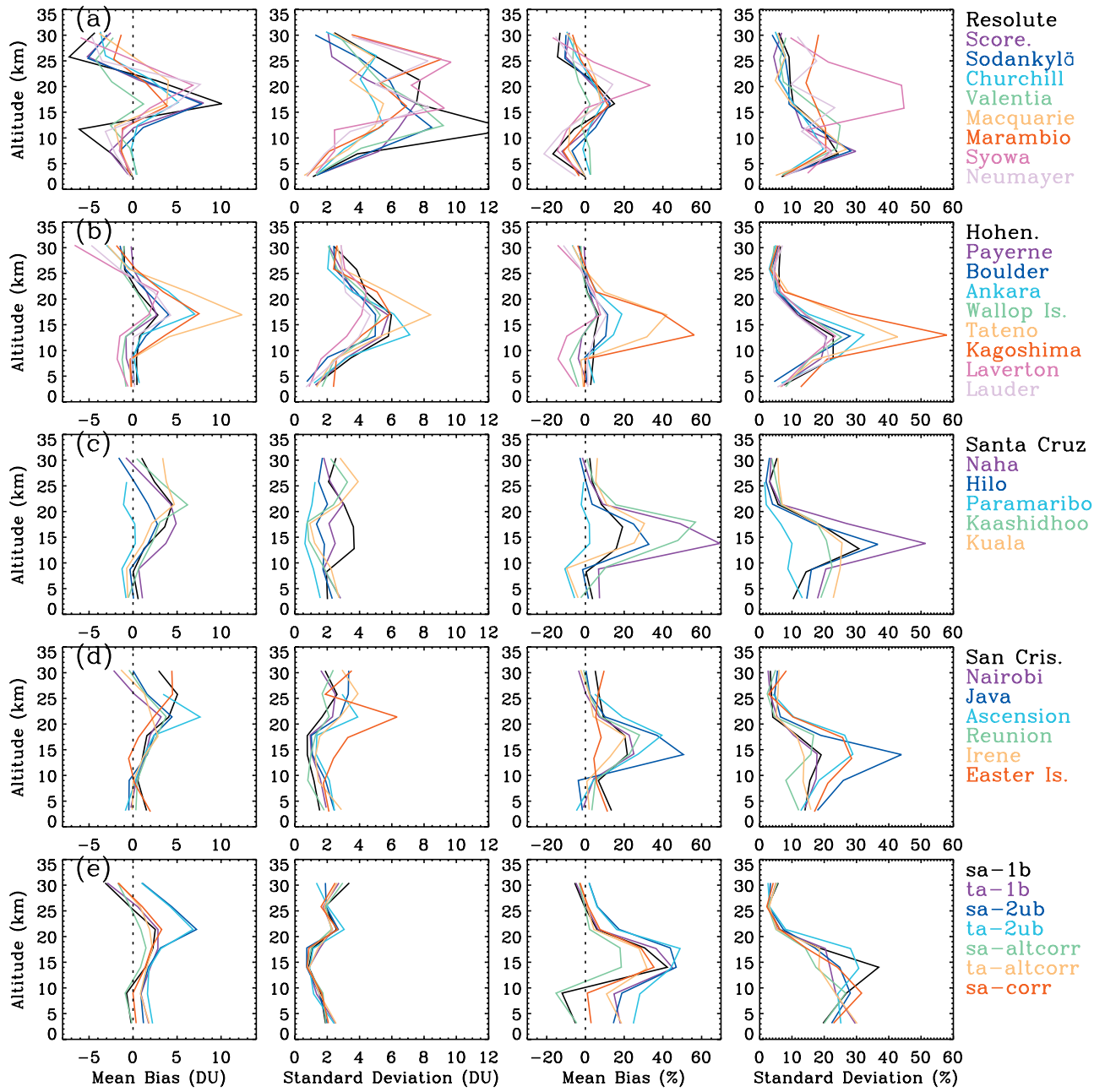

Figure 6. Absolute (columns 1–2) and relative (columns 3–4) mean biases and 1σ standard deviations between GOME and ozonesonde data (GOME minus ozonesonde) at each station, arranged into five groups: (a) >50°N and <50°S south, (b) 30°N/S–50°N/S, (c) 0°–30°N, (4) 30°S–0°S, and (5) American Samoa and Tahiti. Figure 6e shows comparisons for 1% buffer and 2% unbuffered sensor solutions (labeled as "1b" and "2ub"), 1% buffer with altitude correction by shifting the profiles downward by 200 m (labeled as "altcorr"), and homogenized ozonesonde data at America Samoa ("sa-corr").

cantly larger than ozonesonde values in these two layers over these stations.

[24] Figure 6e compares profile biases between 2% ub and 1% b at American Samoa and Tahiti. Before switching the sensor solution, significant positive MBs of 30–45% occur over 10∼20 km for both American Samoa and Tahiti. After switching to 2% ub, the biases generally increase by 5–20%, consistent with the difference between 1% b and 2% ub measurements [*Johnson et al.*, 2002; *Smit and Sträter*, 2004b]. With the altitude-dependent correction, the biases are closer to those with 1% b in both the troposphere and stratosphere. This also supports that systematic biases mainly occur in 2% ub data instead of 1% b.

[25] Ozonesonde measurements have a certain response time. The time for a 90% response to a step change is approximately 50s [*WMO*, 1998], which translates to an altitude registration lag of ∼200 m, which is not corrected in practice [*WMO*, 1998]. Figure 6e also shows that the MBs for 1% b data over ∼10–20 km at American Samoa and Tahiti are reduced to ∼20% and ∼30%, respectively,

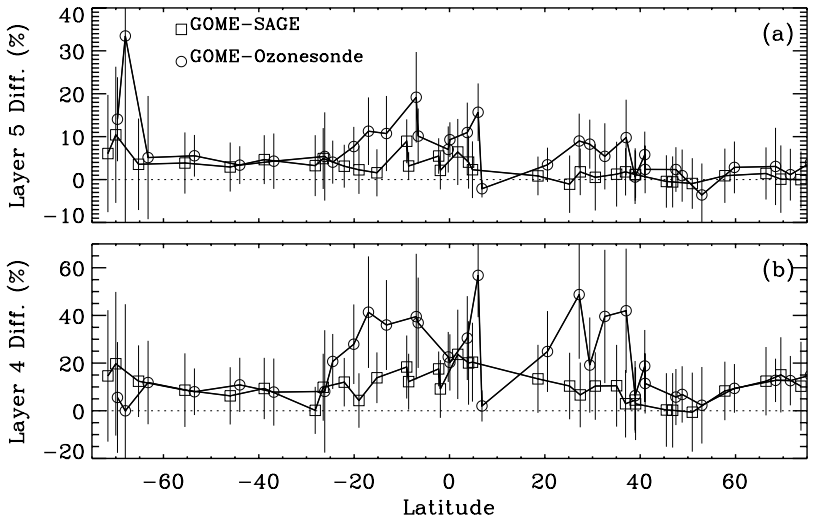

Figure 7. Relative mean biases and 1σ standard deviations for comparisons of column ozone over GOME retrieval layer 4 (∼15–20 km) and 5 (∼20–25 km) between GOME and SAGE II (1996 to May 1998)/ozonesonde (1996–1999). The x axes for squares and circles are shifted by −1° and 1°, respectively.

after shifting down the measurements by 200 m. This correction reduces the biases by 5–15% at layers 3–4 and less than 5% at layer 5 for most of the stations within 40°N–40°S depending on the profile shape. With this correction, the relative MBs over ∼10–20 km are reduced to 10–30% at all the tropical stations except for Kaashidhoo (∼40%). These biases are slightly larger than the GOME/SAGE II biases of 10–20% over ∼15–20 km. The remaining biases of 10–30% (1–4 DU) between GOME, ozonesonde, and SAGE over 10–20 km in the tropics, where ozone concentration is low, could be due to errors in GOME retrievals and SAGE II retrievals (e.g., usually have large uncertainties over these altitude regions). It may also originate from ozonesonde data caused by overremoval of background signals (predetermined before flight) since the background signal may decline with declining pressure, which can be significant in regions of low ozone conditions [*Smit and Kley*, 1998; *WMO*, 1998].

[26] The above intercomparison with GOME retrievals as an intermediate discloses the performance inhomogeneity among the ozonesonde data set (see Figures 3 and 6–7), especially in the stratosphere and upper troposphere over tropical and CI stations. Because the GOME/ozonesonde biases vary from station to station or even vary within a particular station, greatly depending on ozonesonde techniques (e.g., ECC, BM, CI), sensor solutions (e.g., 1% b, 2% ub), instrument model (e.g., EnSci, SPC), and data processing (e.g., normalization method), it is difficult to perform a reliable and consistent validation of satellite observations over these altitude regions without taking the ozonesonde operational characteristics into account. The homogenization procedure used at three CMDL stations, which applies altitude-dependent correction and scales the profiles to concurrent Dobson total ozone with SBUV extrapolation, are shown to greatly improve the data consistency. Thus it is important to homogenize available ozonesonde data sets and standardize future operating procedures of ozonesonde measurements for satellite validations, as have also been recommended in various JOISE activities [*Smit and Kley*, 1998; *Smit and Sträter*, 2004a, 2004b].

4. Summary

[27] We compare SCO and ozone profiles retrieved from Global Ozone Monitoring Experiment (GOME) data with ozonesonde and Stratospheric Aerosol and Gas Experiment II (SAGE II) data during 1996–1999. GOME SCO over the altitude range ∼15–35 km agrees with SAGE SCO to within 2.5 DU (1.5%) on average usually without significant spatiotemporal variation, demonstrating the spatiotemporal consistency of GOME retrievals. Although there are good agreements between GOME and ozonesonde SCO at most middle and high-latitude stations (>30°N/S), GOME SCO is systematically larger than ozonesonde SCO by 8–20 DU at carbon iodine and most tropical stations within 30°N–30°S. The intercomparisons among GOME, SAGE II and ozonesonde data with additional comparisons with TOMS and Dobson total ozone illustrate that those large SCO biases mainly result from ozonesonde underestimates of stratospheric ozone. Some of these large biases can be explained by changes in the strength of sensor solution and whether it is buffered. For example, the switch of sensor solution from 1% KI buffered to 2% KI unbuffered increases the GOME/sonde biases at American Samoa and Tahiti by 11–16 DU during the 1998–1999 period.

[28] The GOME/sonde profile comparison show similar altitude-dependent bias patterns at most stations, indicating systematic errors in the GOME retrievals. These biases are likely due to the residual wavelength-dependent radiometric calibration in the GOME radiance spectra after a second-order polynomial correction. GOME retrievals are significantly larger than ozonesonde observations over the altitude range ∼10–20 km, where ozone concentration is low, for most carbon iodine (30–70%) and tropical (20–55%) stations. However, the GOME/SAGE II biases are usually 10–20% over ∼15–20 km. The uncorrected altitude hys-

teresis in ozonesonde data sets introduces a 5–15% error over 10–20 km. The remaining biases among various measurements under those low ozone conditions may result from errors from both satellite retrievals and ozonesonde measurements (e.g., background signal removal in ozonesonde observations).

[29] The GOME/SAGE II biases around ozonesonde stations usually do not vary much from station to station. However, the GOME/sonde biases show larger variation from station to station especially over carbon iodine and tropical stations, and greatly depend on ozonesonde techniques, instrument type, sensor solution, and the total ozone normalization. The inhomogeneity in ozonesonde performance makes it difficult to perform a reliable and consistent validation of satellite observations over these altitude regions without considering the ozonesonde operational characteristics. The homogenization (applying altitude-dependent correction and scaling to concurrent Dobson total ozone with SBUV extrapolation), used at Boulder, Hilo, and American Samoa stations, is shown to greatly reduce the data inhomogeneity. Therefore it is important to homogenize available ozonesonde data sets and standardize future operating procedures of ozonesonde measurements for reliable satellite validation.

[30] **Acknowledgments.** This study is supported by NASA and the Smithsonian Institution. We thank the WOUDC, SHADOZ, CMDL, and their data originators for providing ozonesonde measurements and the NASA LRC and LRAB for providing the SAGE II data. We also thank S. J. Oltmans, B. J. Johnson, and J. A. Logan for helpful discussion on the comparison with ozonesonde observations. Finally, we acknowledge three anonymous reviewers' constructive comments to improve the paper.

References

Allen, D. R., and R. A. Reck (1997), Daily variations in TOMS total ozone data, *J. Geophys. Res.*, *102*, 13,603–13,608.

Cunnold, D. M., W. P. Chu, R. A. Barnes, M. P. McCormick, and R. E. Veiga (1989), Validation of SAGE II ozone measurements, *J. Geophys. Res.*, *94*, 8447–8460.

Cunnold, D. M., H. Wang, W. P. Chu, and L. Froidevaux (1996), Comparisons between Stratospheric Aerosol and Gas Experiment II and microwave limb sounder ozone measurements and aliasing of SAGE II ozone trends, *J. Geophys. Res.*, *101*, 10,061–10,075.

Cunnold, D. M., H. J. Wang, L. Thomason, J. Zawodny, J. A. Logan, and I. A. Megretskaia (2000), SAGE (v5. 96) ozone trends in the lower stratosphere, *J. Geophys. Res.*, *105*, 4445–4457.

Hasekamp, O. P., and J. Landgraf (2001), Ozone profile retrieval from backscattered ultraviolet radiances: The inverse problem solved by regularization, *J. Geophys. Res.*, *106*, 8077–8088.

Hilsenrath, E., et al. (1986), Results from the Balloon Ozone Intercomparison Campaign (BOIC), *J. Geophys. Res.*, *91*, 13,137–13,152.

Hoogen, R., V. V. Rozanov, and J. P. Burrows (1999), Ozone profiles from GOME satellite data: Algorithm description and first validation, *J. Geophys. Res.*, *104*, 8263–8280.

Johnson, B. J., S. J. Oltmans, H. Vömel, H. G. J. Smit, T. Deshler, and C. Kröger (2002), Electrochemical concentration cell (ECC) ozonesonde pump efficiency measurements and tests on the sensitivity to ozone of buffered and unbuffered ECC sensor cathode solutions, *J. Geophys. Res.*, *107*(D19), 4393, doi:10.1029/2001JD000557.

Komhyr, W. D. (1986), Operations on handbook—Ozone measurements to 40-km altitude with model 4A electrochemical concentration cell (ECC) ozonesondes, *NOAA Tech. Memo. ERLARL-149*, 49 pp.

Labow, G. J., R. D. McPeters, and P. K. Bhartia (2004), A comparison of TOMS & SBUV version 8 total column ozone data with data from ground stations, in *Proceedings of the XX Quadrennial Ozone Symposium*, edited by C. S. Zerefos, pp. 123–124, Int. Ozone Comm., Athens.

Liu, X., K. Chance, C. E. Sioris, R. J. D. Spurr, T. P. Kurosu, R. V. Martin, and M. J. Newchurch (2005), Ozone profile and tropospheric ozone retrievals from the Global Ozone Monitoring Experiment: Algorithm description and validation, *J. Geophys. Res.*, *110*, D20307, doi:10.1029/2005JD006240.

Liu, X., et al. (2006), First directly retrieved global distribution of tropospheric column ozone from GOME: Comparison with the GEOS-CHEM model, *J. Geophys. Res.*, *111*, D02308, doi:10.1029/2005JD006564.

Logan, J. A. (1999), An analysis of ozonesonde data for the troposphere: Recommendations for testing 3-D models and development of a gridded climatology for tropospheric ozone, *J. Geophys. Res.*, *104*, 16,115–116,149.

Logan, J. A., et al. (1999), Trends in the vertical distribution of ozone: A comparison of two analyses of ozonesonde data, *J. Geophys. Res.*, *104*, 26,373–26,399.

Mauldin, L. E., III, N. H. Zaun, M. P. McCormick Jr., J. H. Guy, and W. R. Vaughn (1985), Stratospheric Aerosol and Gas Experiment II instrument: A functional description, *Opt. Eng.*, *24*, 307–317.

Müller, M. D., A. K. Kaifel, M. Weber, S. Tellmann, J. P. Burrows, and D. Loyola (2003), Ozone profile retrieval from Global Ozone Monitoring Experiment (GOME) data using a neural network approach (Neural Network Ozone Retrieval System (NNORSY)), *J. Geophys. Res.*, *108*(D16), 4497, doi:10.1029/2002JD002784.

Munro, R., R. Siddans, W. J. Reburn, and B. Kerridge (1998), Direct measurement of tropospheric ozone from space, *Nature*, *392*, 168–171.

Newchurch, M. J., et al. (2000), Upper-stratospheric ozone trends 1979–1998, *J. Geophys. Res.*, *105*, 14,625–14,636.

Newchurch, M. J., E.-S. Yang, D. M. Cunnold, G. C. Reinsel, J. M. Zawodny, and J. M. Russell (2003), Evidence for slowdown in stratospheric ozone loss: First stage of ozone recovery, *J. Geophys. Res.*, *108*(D16), 4507, doi:10.1029/2003JD003471.

Oltmans, S. J., et al. (2001), Ozone in the Pacific tropical troposphere from ozonesonde observations, *J. Geophys. Res.*, *106*, 32,503–32,525.

Randel, W. J., R. S. Stolarski, D. M. Cunnold, J. A. Logan, M. J. Newchurch, and J. M. Zawodny (1999), Trends in the vertical distribution of ozone, *Science*, *285*, 1689–1692.

Smit, H. G. J., and D. Kley (1998), The 1996 WMO International intercomparison of ozonesondes under quasi flight conditions in the environmental simulation chamber at Jülich, report, World Meteorol. Organ., Geneva, Switzerland.

Smit, H. G. J., and W. Sträter (2004a), JOSIE-1998: Performance of ECC-ozone sondes of SPC-6A and ENSCI-Z type, report, World Meteorol. Organ., Geneva, Switzerland.

Smit, H. G. J., and W. Sträter (2004b), JOSIE-2000: The 2000 WMO international comparison of operating procedures for ECC-ozone sondes at the environmental simulation facility at Jülich, report, World Meteorol. Organ., Geneva, Switzerland.

Thompson, A. M., et al. (2003), Southern Hemisphere Additional Ozonesondes (SHADOZ) 1998–2000 tropical ozone climatology: 1. Comparison with Total Ozone Mapping Spectrometer (TOMS) and ground-based measurements, *J. Geophys. Res.*, *108*(D2), 8238, doi:10.1029/2001JD000967.

van der A, R. J., R. F. van Oss, A. J. M. Piters, J. P. F. Fortuin, Y. J. Meijer, and H. M. Kelder (2002), Ozone profile retrieval from recalibrated Global Ozone Monitoring Experiment data, *J. Geophys. Res.*, *107*(D15), 4239, doi:10.1029/2001JD000696.

Wang, H. J., D. M. Cunnold, and X. Bao (1996), A critical analysis of stratospheric aerosol and gas experiment ozone trends, *J. Geophys. Res.*, *101*, 12,495–12,514.

Wang, H. J., D. M. Cunnold, L. W. Thomason, J. M. Zawodny, and G. E. Bodeker (2002), Assessment of SAGE version 6.1 ozone data quality, *J. Geophys. Res.*, *107*(D23), 4691, doi:10.1029/2002JD002418.

World Meteorological Organization (WMO) (1998), Assessment of trends in the vertical distribution of ozone, *Rep. SPARC/IO3C/GAW*, Geneva, Switzerland.

Wozniak, A. E., J. Fishman, P.-H. Wang, and J. K. Creilson (2005), Distribution of stratospheric column ozone (SCO) determined from satellite observations: Validation of solar backscattered ultraviolet (SBUV) measurements in support of the tropospheric ozone residual (TOR) method, *J. Geophys. Res.*, *110*, D20305, doi:10.1029/2005JD005842.

K. Chance, T. P. Kurosu, X. Liu, and C. E. Sioris, Atomic and Molecular Physics Division, Harvard-Smithsonian Center for Astrophysics, 60 Garden Street, MS-50, Cambridge, MA 02138, USA. (xliu@cfa.harvard.edu)

M. J. Newchurch, Atmospheric Science Department, University of Alabama in Huntsville, 301 Sparkman Drive, Huntsville, AL 35899, USA.