Bias Correction of Multi-sensor Total Column Ozone Satellite Data for 1978–2017

Hiroaki NAOE
Meteorological Research Institute, Tsukuba, Japan

Takanori MATSUMOTO
Japan Meteorological Agency, Tokyo, Japan

Keisuke UENO
Aerological Observatory, Tsukuba, Japan

Takashi MAKI, Makoto DEUSHI
Meteorological Research Institute, Tsukuba, Japan

and

Ayako TAKEUCHI
Japan Meteorological Agency, Tokyo, Japan
Ministry of Land, Infrastructure, Transport and Tourism, Tokyo, Japan

(Manuscript received 21 May 2019, in final form 10 December 2019)

Abstract

This study constructs a merged total column ozone (TCO) dataset using 20 available satellite Level 2 TCO (L2SAT) datasets over 40 years from 1978 to 2017. The individual 20 datasets and the merged TCO dataset are corrected against ground-based Dobson and Brewer spectrophotometer TCO (GD) measurements. Two bias correction methods are used: simple linear regression (SLR) as a function of time and multiple linear regression (MLR) as a function of time, solar zenith angle, and effective ozone temperature. All of the satellite datasets are consistent with GD within ±2–3 %, except for some degraded data from the Total Ozone Mapping Spectrometer/Earth Probe during a period of degraded calibration and from the Ozone Mapping and Profiling Suite (OMPS) provided from NOAA at an early stage of measurements. OMPS data provided from NASA show fairly stable L2SAT–GD differences. The Global Ozone Monitoring Experiment/MetOp-A and -B datasets show abrupt changes of approximately 8 DU coincident with the change of retrieval algorithm. For the TCO merged datasets created by averaging all coincident data located within a grid cell from the 20 satellite-borne TCO datasets, the differences between corrected and uncorrected TCOs by MLR are generally positive at lower latitudes where the bias correction increases TCO because of low effective ozone temperature. In the trend analysis, the difference between corrected and uncorrected TCO trends by MLR shows clear seasonal and latitudinal dependency, whereas such seasonal and latitudinal dependency is lost by SLR. The root mean square difference of L2SAT–GD for the uncorrected merged datasets, 8.6 DU, is reduced to 8.4 DU after correction using SLR and MLR. Therefore, the...
empirically corrected merged TCO datasets that are converted into time-series homogenization with high temporal-resolution are suitable as a data source for trend analyses as well as assimilation for long-term reanalysis.

**Keywords**  total column ozone; bias correction; satellite level 2; merged total ozone

**Citation**  Naoe, H., T. Matsumoto, K. Ueno, T. Maki, M. Deushi, and A. Takeuchi, 2020: Bias correction of multisensor total column ozone satellite data for 1978–2017. *J. Meteor. Soc. Japan*, 98, 353–377, doi:10.2151/jmsj.2020-019.

1. **Introduction**

The increased halogen loading in the stratosphere caused by human activity has led to a decrease in total ozone abundance (e.g., World Meteorological Organization 2019). Changes in the ozone layer are of major concern to scientists and the general public, and thus atmospheric ozone is being closely monitored to assess the recovery of the stratospheric ozone layer from chemical ozone depletion. In response to the observed ozone depletion, the Montreal Protocol on Substances that Deplete the Ozone Layer was agreed upon in 1987, and the successful implementation of the protocol and its later amendments are expected to result in ozone recovery to pre-1980s levels (World Meteorological Organization 2019). Newchurch et al. (2003) performed a statistical analysis of satellite measurements that stratospheric ozone depletion, which is controlled mainly by gas-phase catalytic cycles, slowed globally after 1997. However, the estimated dates of return of the total column ozone (TCO) to pre-1980 values for the baseline scenario (RCP-6.0) are different for the mid-latitudes in the northern and southern hemispheres (NH and SH) and significantly delayed for the Antarctic ozone. They will probably occur before the mid-century (~2035) around the mid-century for annually averaged NH/SH mid-latitude ozone, and after the mid-century (~2060) for the Antarctic ozone hole (World Meteorological Organization 2019).

Global-scale satellite observations of TCO have been performed since the early 1970s, and regular and continuous ozone monitoring by the Total Ozone Mapping Spectrometer (TOMS) and Solar Backscatter Ultraviolet (SBUV) instruments onboard the Nimbus-7 satellite started in 1978 (McPeters et al. 1996; Bhartia et al. 2013). Since then, ozone observations by multiple instruments on different polar-orbiting platforms have been available. Some of these instruments are the Global Ozone Monitoring Experiment (GOME) onboard the European Remote Satellite-2 (ERS-2) (Bodeker et al. 2001), the Ozone Monitoring Instrument (OMI) onboard the Earth Observation System Aura satellite (Koukouli et al. 2012), and the Ozone Mapping and Profiling Suite (OMPS) onboard the Suomi National Polar-orbiting Partnership (S-NPP) satellite (Flynn et al. 2014).

Satellite-borne instruments have an important advantage for ozone measurements compared to sparsely distributed ground-based networks because they can provide daily global TCO maps with a resolution that is sufficient to detect meteorological variability across regions. However, they suffer from several disadvantages, including instrument calibration drift and occasional gaps in data continuity due to changes in retrieval algorithms and observation conditions. In contrast, ground-based instruments are regularly cross-calibrated and provide continuous time series data, but they are spatially scattered, and their coverage of the globe is far from complete.

So far, some differences within ±3% were reported between the TCO measurements of TOMS or GOME instruments and ground-based observations (Fioletov et al. 2008). Antón et al. (2010) evaluated TOMS/Earth Probe (EP) TCO data and concluded that TOMS data for the period of 2000–2004 should not be used for trend analysis due to degraded calibration issues. Van Roozendael et al. (2012) reported an updated GOME Data Processor version 5 (GDP5) spectral fitting algorithm for reprocessing 16-year GOME data records and they indicated that the updated GDP5 showed a clear improvement with reduced solar zenith angle and seasonal dependences for nearly all regions compared with ground-based measurements. Loyola et al. (2011) evaluated GOME-2/MetOp-A TCO measurements for 3 years (2007–2009) using GDP version 4.4 retrieval algorithm and they found that GOME-2 TCO slightly underestimated (overestimated) ground-based ozone by 0.5–1% (~0.5%) over the mid-latitudes of the NH (SH). Meanwhile, Kroon et al. (2011) evaluated OMI ozone profiles against co-flying instruments and collocated ozone profiles and they showed an agreement of stratospheric ozone profiles within 20% with global correlate data, with
a systematic positive bias of ~ 60 % in the tropics and 30 % in the mid-latitude regions. Furthermore, TCO datasets of the European sensors GOME, SCanning Imaging Absorption SpectroMeter for Atmospheric CartographHY (SCIAMACHY), and GOME-2 were reprocessed with GOME-type Direct FITting (GODFIT) version 3 retrieve algorithm (Lerot et al. 2014; Koukouli et al. 2015). The ozone datasets had good quality and less sensitivity to instrumental degradation owing to a new reflectance soft-calibration scheme. These studies suggest that in order to obtain accurate global ozone records, satellite measurements must be validated against measurements made by ground-based instruments. Therefore, a merged ozone dataset from multiple satellite-acquired ozone after the validation is suitable for trend analysis and climate change studies.

One of the major difficulties in assessing long-term global ozone variations is data inhomogeneity (Fioletov et al. 2002). Satellite measurements are obtained by different instruments and retrieved by different algorithms; furthermore, the instruments are subject to problems such as radiation damage (van der A et al. 2010). Although SBUV instruments are quite similar in design, different orbital characteristics between the satellites resulted in measurements made at different local times and solar zenith angles (Miller et al. 2002; Frith et al. 2014). Thus, the drifting orbits for the series of SBUV/2 instruments on NOAA satellites are possibly the biggest source of uncertainty regarding the determination of long-term ozone change from these instruments. Operational satellite changes and observation interruptions result in systematic errors for datasets that change over time, and these errors must be taken into account when estimating the ozone recovery trend.

To date, several merged ozone datasets have been created. For example, Bodeker et al. (2001) constructed a merged ozone dataset from TOMS and GOME datasets after removing latitudinal and temporal drift and biases by intersatellite comparisons and by examining the differences between satellite overpass and ground-based measurements as a function of time and latitude. Meanwhile, Stolarski and Frith (2006) constructed a merged ozone dataset from TOMS and SBUV version 8 retrieved data by intersatellite comparisons and by averaging the data during instrument overlapped periods. Besides, merged monthly zonal mean version 8.6 ozone data were constructed from nine independent SBUV-type instruments operating in the 8 am–4 pm equator crossing time range (McPeters et al. 2013; Frith et al. 2014). These merged datasets validated by intersatellite comparison or overpass ground-based datasets are used for trend analysis based on monthly data.

On the other hand, ozone datasets are used for long-term reanalysis such as the European Centre for Medium-Range Weather Forecasts Reanalysis-Interim (ERA-Interim) (Dragani 2011), the Japanese 55-year Reanalysis conducted by the Japan Meteorological Agency (JRA-55) (Kobayashi et al. 2015), and Modern-Era Retrospective Analysis for Research and Applications version 2 (MERRA-2) (Wargan et al. 2017). However, these ozone datasets are not homogenized for the usage of the reanalysis. In contrast, van der A et al. (2010, 2015) constructed merged TCO datasets from the assimilation based on high temporal-resolution satellite data in which the differences between Level 2 (L2) retrieval datasets and ground-based observations were corrected for systematic biases as a function of solar zenith angle (SZA), viewing angle, time, and effective ozone temperature ($T_{eff}$).

In this study, we apply the method of van der A et al. (2010, 2015) for the bias correction of 20 satellite datasets comprising the latest TCO retrievals, to produce continuous and homogeneous TCO datasets covering 40 years from 1978 to 2017. We first evaluate biases in the individual satellite TCO datasets by comparing the satellite data with spatially and temporally collocated ground-based Dobson and Brewer spectrophotometer networks. Then, we create individual corrected datasets by weighted averaging all data within each grid cell. The empirically corrected merged TCO dataset with high temporal resolution (e.g., daily or hourly) constructed in this manner is thus suitable for trend analysis as well as assimilation of long-term reanalysis. Besides, the corrected datasets are useful for the post-processing of Dobson spectrophotometer data, because the retrieval algorithm for the Dobson data does not account for natural interannual variations in stratospheric temperature, which significantly affects ozone absorption coefficients (Koukouli et al. 2016).

This paper is organized as follows. In Section 2, we describe the L2 satellite-borne TCO datasets used, all of which are available from public archives. Section 3 presents our methods, including two bias correction methods, and methods used to calculate the differences between coincident satellite overpass data and ground-based observations and to produce Level 3 (L3) TCO datasets. Section 4 presents the corrected TCO data for biases by simple linear regression (SLR) as a function of time, and Section 5 presents the corrected data by multiple linear regression (MLR) as a function of time, SZA, and $T_{eff}$. Section 6 investi-
gates the TCO trend in the corrected and uncorrected merged TCO datasets, and Section 7 presents our concluding remarks.

2. Datasets

In this study, 20 L2 satellite TCO datasets covering 40 years from 1978 to 2017 are acquired from public archives (Table 1). The short names used to refer to these datasets are mainly based on the name of collecting instruments. For datasets composed of data acquired by the same type of instrument but on a different satellite or by a different retrieval algorithm, the dataset name includes the name of the satellite as well as algorithm or provider. Thus, the 20 datasets used are: TOMS/N7, TOMS/EP, GOME, TOGOMI, SCIAMACHY, TOMOSI, OMI-TOMS, OMI-DOAS, GOME-2A, GOME-2B, OMPS-NASA, OMPS-NOAA, SBUV/N7, SBUV/N9, SBUV/N11, SBUV/N14, SBUV/N16, SBUV/N17, SBUV/N18, and SBUV/N19. The instruments that acquired the ozone data and the datasets are briefly described below.

2.1 Ground-based data

Ground-based Dobson and Brewer spectrophotometer TCO measurements are collected from “Total Ozone” of the archive of the World Ozone and Ultraviolet Data Center (WOUDC). Daily TCO values, either the best representative values or daily averages, are usually submitted to WOUDC by a contributor. For example, the Japan Meteorological Agency has submitted the best representative values of Dobson observations and the daily averages of Brewer observations. By visually inspecting the differences between satellite overpass and ground-based TCO measurements as a function of time for each station, data with low number of observations, large data gap, sudden jumps, and very large offsets are rejected. Thus, 34 Brewer and 39 Dobson TCO measurement stations are selected in this study (Tables A1, A2, respectively). Single monochromator Brewer (and Dobson) spectrophotometers are susceptible to scattered light error for large SZA observations, whereas double-monochromator Brewer spectrophotometers

| Name*, Algorithm Version | Time Period | References |
|-------------------------|-------------|------------|
| TOMS/N7, TOMS v8        | Nov 1978 – May 1993 | TOMS Science Team (1996)¹ |
| TOMS/EP, TOMS v8        | Sep 1996 – Aug 2003 | TOMS Science Team (1998)² |
| GOME, GODFIT v4         | Jun 1996 – Jun 2011 | Ozone CCI (2012)³ |
| SCIAMACHY, GODFIT v4    | Aug 2002 – Apr 2012 | Ozone CCI (2012)³ |
| TOGOMI, TOGOMI v2.1     | Jun 1995 – Jul 2011 | Valks et al. (2004), TEMIS (2005)***⁴ |
| TOSOMI, TOSOMI v2.0     | Aug 2002 – Apr 2012 | Eskes et al. (2005), van der A et al. (2006)⁵ |
| OMI-TOMS, TOMS v8       | Oct 2004 – Dec 2017 | Bhartia (2005)⁶ |
| OMI-DOAS, DOAS v1       | Oct 2004 – Dec 2017 | Veefkind (2006)⁷ |
| GOME-2A, GDP v4.6-8**   | Jan 2007 – Dec 2017 | DLR (2012)****⁸ |
| GOME-2B, GDP v4.7-8**   | Jan 2013 – Dec 2017 | DLR (2012)****⁸ |
| OMPS-NOAA, TOMS v7      | Mar 2012 – Dec 2017 | Flynn et al. (2014) |
| OMPS-NASA, TOMS v8      | Nov 2011 – Dec 2017 | Jaross (2017)¹⁰ |
| SBUV/N7, SBUV v8.6      | Oct 1978 – Jun 1990 | Frith et al. (2014)¹¹ |
| SBUV/N9, SBUV v8.6      | Feb 1985 – Feb 1998 | Frith et al. (2014)¹¹ |
| SBUV/N11, SBUV v8.6     | Dec 1988 – Mar 2001 | Frith et al. (2014)¹¹ |
| SBUV/N14, SBUV v8.6     | Feb 1995 – Sep 2006 | Frith et al. (2014)¹¹ |
| SBUV/N16, SBUV v8.6     | Oct 2000 – Jun 2014 | Frith et al. (2014)¹¹ |
| SBUV/N17, SBUV v8.6     | Jul 2002 – Jan 2013 | Frith et al. (2014)¹¹ |
| SBUV/N18, SBUV v8.6     | Jun 2005 – Dec 2012 | Frith et al. (2014)¹¹ |
| SBUV/N19, SBUV v8.6     | Feb 2009 – Jun 2016 | Frith et al. (2014)¹¹ |

* The names of datasets are generally based on the instruments used to acquire the data. Exceptions are TOGOMI and TOSOMI, which were acquired by GOME and SCIAMACHY instruments, respectively. The SBUV/N9, /N11, /N14, /N16, /N17, /N18, and /N19 datasets were acquired by the second generation (SBUV/2) instruments. Both GOME-2A and -2B instruments were acquired by the GOME-2 instrument. OMPS products were acquired by the OMPS Nadir Mapper (NM).

** January 2007 to July 2013, GDP 4.6; July 2013 to October 2016, GDP 4.7; October 2016 onwards, GDP 4.8.

*** Tropospheric Emission Monitoring Internet Service (2005).

**** Deutsches Zentrum für Luft und Raumfahrt (2012), hereafter referred to DLR (2012).

¹–¹¹ Accessed on ²⁶ October 2017, ⁷ March 2016, ⁷ August 2015, ⁴¹ December 2015, ²⁵ March 2016, ¹⁴ February 2017, ³¹ January 2017, ⁴¹ February 2018, ⁶ February 2018, ⁷ August 2017, and ¹⁰ December 2015.
enable more precise observations (Karppinen et al. 2015). Thus, single or/and double Brewer measurements are listed in Table A1 for reference.

2.2 TOMS

The TOMS onboard Nimbus 7 (TOMS/N7) with TCO version 8 retrieval algorithm provided the daily TCO global distribution (TOMS Science Team 1996; McPeters et al. 1996). TOMS/N7 had an orbit altitude of 935 km and this instrument performed cross-track scanning with a 102° field of view (FOV) using 35 sample points at 3° intervals, corresponding to a 50 km resolution at the Earth’s surface for near-nadir angle observations. This basic retrieval algorithm uses two wavelengths to derive TCO. The shorter wavelength, which has a stronger continuum ozone absorption in the Huggins band (310 – 340 nm), is sensitive to TCO. Meanwhile, the longer wavelength, which has weak ozone absorption, characterizes the underlying surface and clouds.

The TOMS onboard the EP satellite (TOMS/EP) started measurements from an initial orbit of 500 km with an inclination of 98°. In December 1997, the orbit was elevated to an altitude of 739 km in order to increase the daily global coverage to 90 % (TOMS Science Team 1998; McPeters et al. 1998). The TOMS/EP TCO measurements were processed with TCO version 8 retrieval algorithm. After the first few years of observation, TOMS/EP had serious optical degradation and the calibration had to be empirically corrected. Thus, it is not recommended to use degraded data from EP for trend analysis (Antón et al. 2010).

TOMS instruments were also mounted on the Meteor-3 satellite of the former USSR and the Japanese ADvanced Earth Observing Satellite (ADEOS). The Meteor-3 satellite had three-diffuser systems (cover, working, and reference sequence) for backscatter ultraviolet instruments. It had a non-synchronous near-circular and near-polar prograde orbit that slowly drifted with local time. The reference diffuser was usually exposed for one sequence every 106 days (half the period of orbital precession). This precession posed a serious problem for trend determination from the TOMS/Meteor-3 instrument. The ADEOS spacecraft failed after only about a year in orbit; hence, little data are available. TOMS/Meteor-3 and TOMS/ADEOS measurements are not used in this study because their L2 data were not available from public archives.

2.3 SBUV

The SBUV instrument onboard Nimbus-7 (SBUV/N7) measured backscattered ultraviolet radiation at 12 different wavelengths to determine TCO and vertical ozone profiles (Bhartia et al. 2013). Datasets SBUV/N9, /N11, /N14, /N16, /N17, /N18, and /N19 comprise ozone measurements acquired by second-generation instruments (SBUV/2) onboard NOAA series satellites (NOAA-9, -11, -14, -16, -17, -18, and -19, respectively) (Frith et al. 2014). Version 8.6 retrieval algorithm was used to reprocess the SBUV TCO measurements (Bhartia et al. 2013; Frith et al. 2014). Both SBUV and SBUV/2 were nadir-viewing instruments with an FOV of approximately 11.3° × 11.3°, which corresponded to an area of approximately 180 km × 180 km (168 km × 168 km) for SBUV (SBUV/2) on the Earth’s surface. The acquisition of one set of SBUV measurements, only during the daylight portion of each orbit, took 32 s. About 14 orbits were covered per day, with a 26° separation at the equator.

2.4 GOME

GOME, onboard ERS-2, was a nadir-viewing scanning spectrometer designed to measure a range of atmospheric trace constituents in the troposphere and stratosphere (Deutsches Zentrum für Luft und Raumfahrt (DLR) 2012). The largest FOV of GOME was 320 × 40 km, and three cross-track measurements provided global coverage at the equator every 3 days. An operational TCO dataset was generated by GDP version 4, using a Differential Optical Absorption Spectroscopy (DOAS) algorithm (van Roozendael et al. 2006). A DOAS-type algorithm compares the spectral fitting of effective slant columns and air-mass factor computations (mainly from look-up tables). In accordance with the Ozone Climate Change Initiative (CCI) project of the European Space Agency (ESA) (Lerot et al. 2014), the TCO dataset was reprocessed using GODFIT algorithm version 4 (Lerot et al. 2010; Van Roozendael et al. 2012), and this updated reprocessed dataset is used in this study.

Another GOME nadir TCO dataset was retrieved using the Total Ozone retrieval scheme for GOME developed for the OMI instrument (TOGOMI version 2.1) (Valks et al. 2004; Tropospheric Emission Monitoring Internet Service 2005).

2.5 SCIAMACHY

The SCanning Imaging Absorption spectroMeter for Atmospheric CHartographY (SCIAMACHY) onboard the ESA Environmental satellite (Envisat) had a nadir mode that detected upwelling solar radiation scattered in the atmosphere and reflected by the Earth’s surface. It had a 30 × 60 km along-track FOV and a swath
width of 960 km (Ozone CCI 2012; Ebojie et al. 2014; Lerot et al. 2014). An operational TCO dataset was generated using SCIAMACHY Data Processor (SDP) versions 3–5 by DOAS technique (Lerot et al. 2009). The SCIAMACHY TCO dataset was also reprocessed using GODFIT algorithm to achieve a high level of consistency with other instruments (Lerot et al. 2014; Van Roozendael et al. 2012). This updated reprocessed dataset is used in this study.

Another SCIAMACHY TCO dataset was retrieved using the Total Ozone retrieval scheme for SCIAMACHY, which used OMI-DOAS algorithm (TOSOMI version 2.0) (Eskes et al. 2005; van der A and Eskes 2006), which was an application of TOGOMI algorithm to SCIAMACHY.

2.6 OMI

OMI is a Dutch-Finnish nadir-viewing near-UV/visible charge-coupled device (CCD) spectrometer onboard NASA's Earth Observation System's Aura satellite (Bhartia 2002; OMI team 2012). OMI measurements cover the 264–504 nm spectra with a spectral resolution of 0.42–0.63 nm. The nominal surface FOV is 13 × 24 km at nadir with a swath width of 2600 km. Typically, 60 cross-track measurements are made per swath, and 1650 swath observations are made per orbit. The TCO measurements are processed using TOMS version 8 algorithm, which uses two wavelengths (317.5 nm and 331.2 nm) to obtain the TCO measurements. The other four wavelengths are used for diagnostics and error correction (Bhartia 2005).

Another TCO dataset is derived using the DOAS-based TCO algorithm of the Royal Netherlands Meteorological Institute (KNMI) (Veefkind 2006), which was originally used to process data from the GOME instrument onboard ERS-2. Because this algorithm is most suitable for weakly absorbing trace gases, it estimates TCO data from longer wavelengths than those used by the TOMS algorithm.

2.7 GOME-2A and -2B

The GOME-2 instruments, onboard the ESA Operational Meteorological Satellites (MetOp-A and -B; datasets GOME-2A and GOME-2B), have a main-channel FOV of 80 × 40 km and obtain 24 cross-track measurements on a swath with a nominal width of 1920 km (DLR 2012). From January 2007 to 15 July 2013, GOME-2A data were processed with GDP 4.6, and from 15 July 2013 to 15 October 2016, GOME-2A and -2B data were processed with GDP 4.7. GDP 4.8 has been used to process all data since 15 October 2016 (http://o3msaf.fmi.fi/products/oto_o3.html).

2.8 OMPS

The OMPS Nadir Mapper (NM) onboard the S-NPP satellite is a TCO sensor operated by the Joint Polar Satellite System (JPSS) program (Godin 2014). OMPS-NM measurements are processed in near real-time at the NOAA Interface Data Processing Segment using the Environmental Data Record algorithm, which is similar to NASA’s version 7 TCO retrieval algorithm (Flynn et al. 2014). OMPS-NM has a 110° cross-track FOV with 35 cross-track bins and a 0.27° along-track slit width, corresponding to a 50-km resolution at the Earth’s surface. It provides instantaneous coverage of a 2800-km-wide swath at the Earth’s surface. In this study, the OMPS-NOAA dataset comprises OMPS data from the JPSS NOAA site.

Another OMPS-NM TCO dataset is provided by NASA’s Goddard Earth Science Data and Information Services Center, which uses an algorithm that is similar to NASA TOMS version 8 retrieval algorithm used for OMI (Jaross 2017; Kramarova et al. 2014). Because the CCD readout in the NASA product is split at the center, there are typically 36 cross-track measurements per swath and 400 swath observations per orbit. The OMPS-NASA dataset in this study comprises OMPS data from the NASA site.

3. Methods

All L2 binary and ASCII data are first transformed into a uniform ASCII format. Then, uncorrected L3 TCO datasets are created from the uniform L2 ASCII data. In this study, two bias correction methods are used. Biases of datasets are corrected by SLR as a function of time and by MLR as a function of time, SZA, and $T_{eff}$. The systematic mean biases and drift of the individual datasets are evaluated from the differences between coincident L2 satellite overpass data and selected ground-based observations. Meanwhile, $T_{eff}$ is calculated if a L2 retrievals are not available. Then, the linear regression coefficients are used to create corrected L3 TCO datasets from the uniform L2 ASCII data. Finally, we describe a method for estimating the uncertainties of systematic mean biases and drift.

3.1 Uniform ASCII format

All datasets in binary NetCDF, HDF, HDF-EOS5 formats, or original ASCII formats are converted to our uniform ASCII format. Elements of the uniform format are observation time, center location (latitude and longitude) of a satellite pixel, four corner loca-
tions of the pixel, TCO and TCO error, SZA, viewing zenith angle (VZA), \( T_{gf} \), surface flags (sea/land/ice), and error flags. The SBUV datasets contain one file for each observation day while the other datasets contain one file for each orbit. Ground-based observation data from WOUDC in a CSV format are converted to the uniform format.

3.2 Differences between satellite overpass TCO observations and ground-based measurement data (SAT–GD_{diff})

The satellite overpass at a ground-based station is defined when the distance between the center of the satellite pixel and the ground-based station location is less than 50–200 km, depending on the satellite pixel size. Because daily data for Brewer and Dobson measurements are derived from “TotalOzone” of the WOUDC, we assume that ground-based measurements would be at local noon. Thus, we consider a time window within ±4 h (i.e., from 8 am to 4 pm) for satellite overpass. The local noon is estimated from the longitude of the ground-based station. For spatiotemporal collocated satellite overpass and ground-based data, differences between coincident satellite L2 TCO observations (L2SAT) and daily TCO ground-based measurement data (GD) from the Brewer and Dobson spectrophotometers (Table A1 and A2, respectively), SAT–GD_{diff} = L2SAT – GD, are calculated.

In this paper, several criteria are adopted to SZA and cloud amount based on the purpose of the analysis. To analyze the dependence of SZA, cloud amount, and \( T_{gf} \) on SAT–GD_{diff}, satellite overpass data when SZA < 89° and all the cloud amount are used. For the regression analysis, only satellite overpass data at SZA < 80° and cloud amount < 0.8 are used, since SAT–GD_{diff} becomes large at large SZA, and the root mean square of SAT–GD_{diff} (RMSD) are large when cloud amount is close to 1. To create TCO products at each grid mesh, L2SAT at SZA < 84° are used because these data are often flagged as “good”. The cloud amount criterion is set at 0.8.

3.3 Effective ozone temperature (\( T_{ef} \))

The ozone absorption coefficient, which is used by retrieval algorithms to estimate TCO, is strongly dependent on the stratospheric temperature that changes with season and latitude. \( T_{ef} \) is defined, following Koukouli et al. (2016), as the integral of the temperature weighted with ozone profile over altitude as follows:

\[
T_{ef} = \int T(z) O_3(z) dz / \int O_3(z) dz.
\]

In this study, the biases of satellite data as a function of \( T_{ef} \) are corrected as part of MLR fitting. However, the retrieval \( T_{gf} \) is not available in some datasets. In that case, \( T_{ef} \) is computed from JRA-55 (Kobayashi et al. 2015) temperature profile data, \( T(z) \) and the climatological monthly-mean ozone profile, \( O_3(z) \) (Fortuin and Kelder 1998). The temperature data are extracted from the 6-hourly three-dimensional temperature distribution in JRA-55, and \( T(z) \) is derived at local noon from the reanalysis temperature for each ground station by interpolation (van der A et al. 2010).

Temperature correction (van der A et al. 2010; Koukouli et al. 2016) is applied to post-process the Dobson TCO as follows:

\[
TCO_{corr} = TCO_{Dobson} \left\{ 1 - 0.0013(T_{ef} - 226.7) \right\},
\]

where \( TCO_{corr} \) is the corrected Dobson TCO generated using \( T_{ef} \). \( TCO_{Dobson} \) is the Dobson measurement corresponding to the Dobson reference effective temperature of 226.7 K.

3.4 Bias correction

Biases of the individual and merged satellite TCO datasets are corrected in comparison with selected GD by two methods: SLR and MLR. In the first bias correction method, daily SAT–GD_{diff} that is within three times the standard deviation at each ground-based station is averaged over all the stations. Then, to correct the mean bias (offset) and the trend (long-term drift) of the data, SLR is applied to daily station-averaged SAT–GD_{diff} as a function of time. Some datasets contain degraded data because of instrumental problems or because they were obtained during the early stage of the satellite deployment (Antón et al. 2010; Flynn et al. 2014). In such cases, the satellite measurement period is divided into two parts, and linear regression analysis is applied separately to the data collected during each sub-period. Moreover, SBUV datasets of satellites that drifted from an afternoon (ascending) orbit to a morning (descending) orbit are analyzed as different time series. We only use SBUV measurements in the 8 am–4 pm equator crossing time range, with a single exception for SBUV/N11 in the 1994–1995 period to avoid a gap in the time series (Frith et al. 2014).

In the second bias correction method, MLR is applied to daily SAT–GD_{diff} as a function of time, SZA,
and $T_{eff}$ at all the stations rather than daily station-averaged differences.

3.5 Level 3 datasets

We produce hourly L3 TCO datasets by spatially re-sampling L2SAT data to a linear Gaussian grid with TL159 resolution (320 longitudinal intervals × 160 latitudinal intervals). The L2 longitude and latitude data are first converted to 0.1° intervals and then re-mapped to a grid location. Besides, the start and end times of each hourly interval recorded in advance as line numbers of an orbit-based-text file and number of records, are used for the L3 production.

Here, only satellite TCO data collected on the sunlit side of Earth are used. Data flagged as “bad” as well as those that do not satisfy certain criteria (i.e., data collected when SZA > 84° (McPeters et al. 1996), cloud cover > 0.8, TCO values outside the 100–700 DU range) are not used. Basically, for non-SBUV datasets, all data flagged as “good” are used. In the case of SBUV datasets, however, global data coverage during 1993–96 and 2003–04 would be extremely low if all data flagged as “bad” were excluded. To prevent data gaps, we disregarded quality flags and include all SBUV datasets for the L3 production.

Each gridded product in the dataset is generated by first multiplying data within a grid cell by a weighting factor that is inversely proportional to the TCO error, and then averaging the weighted data within the grid cell. In other words, the data with low errors within a grid cell are assumed to be more reliable. In some datasets, satellite FOVs at higher latitudes are larger than the grid cell size. In that case, the average grid cell value is patched to adjacent grid cells, where the number of adjacent grid cells is proportional to the ratio of the FOV area, calculated using spherical trigonometry, to the grid cell area. The daily and monthly L3 datasets are created by summing hourly L3 values.

Moreover, merged datasets are produced in a similar but separate manner by averaging the data located within each grid cell over all datasets multiplied by the weighting factor that is inversely proportional to the TCO error.

3.6 Uncertainties

Data averaging and the uncertainty in the measurements are outlined in Damadeo et al. (2014) and Petropavlovskikh et al. (2019). The uncertainty in the mean is representative of both the uncertainty in the measurements and the overall variance (Damadeo et al. 2014). A simple lag-1 autocorrelation correction is considered for the SLR model that is applied to daily station-averaged SAT–GD$_{diff}$. Since we have an approach for using the station-averaged SAT–GD$_{diff}$, these differences are assumed to be homoscedastic. No seasonal cycle is included in this regression because Fourier components representing the seasonal cycle are relatively small.

The trend analysis is conducted based on two different periods to describe the ozone decrease in the 1980s and 1990s, and the slow ozone increase since the early 2000s (Petropavlovskikh et al. 2019). Fourier components representing 12, 6, 4, and 3 months periodicity are included in the SLR, and other proxies such as the quasi-biennial oscillation, the 11-year solar cycle, etc., are not included.

4. Bias correction as a function of time by simple linear regression (SLR)

In this section, the satellite overpass retrievals of each dataset are compared with coincident GD made by Dobson and Brewer spectrophotometers. The systematic mean bias and drift of SAT–GD$_{diff}$ over time are assessed using SLR. The daily L2SAT is compared with corresponding daily GD recorded at each ground-based station as follows: First, daily SAT–GD$_{diff}$ is calculated for each site, and then the mean daily differences for all the sites are calculated. Second, SLR is applied to the daily mean differences as a function of time to determine the bias (offset) and trend (slope) of each dataset (Table 2).

Figure 1 shows a 31-day-running mean of SAT–GD$_{diff}$ and the regression lines (shown in red) in the 20 datasets. Overall, all the datasets agree with GD within ±2–3 % (< 10 DU), except for periods when the measurements were significantly degraded (TOMS/EP during 2000–2003, and OMPS-NOAA during 2012–2014). In many datasets, the slope (trend) of SAT–GD$_{diff}$ is within ±1 % (< 4 DU) per decade. Exceptions are TOMS/EP during 2000–2003 due to instrument degradation (Antón et al. 2010), and OMPS-NOAA and GOME-2B because of unstable satellite measurements during the early deployment stage. After their early stage, the drift in these two acquired datasets is reduced (i.e., the slope is smaller). TOMS/N7 shows good agreement with GD, with a mean bias of +3 DU (1 %) and a drift of −2 DU per decade, and TOMS/EP shows small deviations of ~ 0.5 % during 1996–1999. However, after 2000, TOMS/EP observations display significant deviations from GD, reaching low values of −12 DU in 2003 with an abnormally low trend of −24 DU per decade. Thus, we exclude the TOMS/EP data from 2000 to 2003 from our trend evaluation, based on the recom-
Recommendations of Antón et al. (2010) and McPeters et al. (2008).

GOME and SCIAMACHY data that were processed with the GODFIT algorithm (Lerot et al. 2010) show almost no slope (within ±0.1 DU per decade) and mean biases of 2.9 and 0.3 DU, respectively (Table 2), compared with GD. The 31-day-running mean peak-to-peak amplitude ranges ~ 6 DU in GOME and 3 DU in SCIAMACHY (Figs. 1c, e, respectively). Meanwhile, the TOGOMI and TOSOMI are similar to those of GOME and SCIAMACHY, respectively, except that the offset in TOGOMI (0.5 %) is reduced in comparison with that in GOME (1 %), and the RMSD is 9.5 DU for TOSOMI compared with 8.7 DU for SCIAMACHY.

OMI-TOMS and OMI-DOAS datasets agree well with GD (Kroon et al. 2008). The time series of SAT–GD_{diff} (Figs. 1g, h, respectively) are almost flat throughout the year, both datasets have a mean bias of −3.1 DU and −1.8 DU, respectively. OMI-TOMS has a slightly larger drift and a smaller RMSD compared with OMI-DOAS. Similar to GOME and SCIAMACHY that were also processed with the DOAS-type algorithm, OMI-DOAS displays seasonal behavior with relatively smaller seasonal amplitude.

The GOME-2A and -2B datasets exhibit seasonal peak-to-peak SAT–GD_{diff} of 2–3 %, owing to using the DOAS algorithm. Their time series show abrupt

| Name          | Time Period         | RMSD1 DU | RMSD2 DU | Offset DU | Trend DU decade⁻¹ |
|---------------|---------------------|----------|----------|-----------|-------------------|
| TOMS/N7       | Nov 1978 – May 1993 | 10.4     | 10.0     | 3.1 ± 0.1 | −2.2 ± 0.4        |
| TOMS/EP       | Jul 1996 – Dec 1999 | 9.4      | 9.3      | −1.5 ± 0.2 | −2.7 ± 2.5        |
|               | Jan 2000 – Aug 2003 | 11.6     | 10.3     | −4.6 ± 0.2 | −24 ± 7           |
| GOME          | Jun 1995 – Jun 2011 | 11.3     | 10.9     | 2.9 ± 0.1 | −0.1 ± 0.4        |
| SCIAMACHY     | Aug 2002 – Apr 2012 | 8.7      | 8.7      | 0.3 ± 0.1 | 0.0 ± 0.6         |
| TOGOMI        | Jun 1995 – Jul 2011 | 10.7     | 10.6     | 1.7 ± 0.1 | −0.9 ± 0.4        |
| TOSOMI        | Aug 2002 – Apr 2012 | 9.5      | 9.5      | 0.0 ± 0.2 | −1.2 ± 1.1        |
| OMI-TOMS      | Oct 2004 – Dec 2017 | 8.8      | 8.1      | −3.1 ± 0.1 | 1.8 ± 0.3         |
| OMI-DOAS      | Oct 2004 – Dec 2017 | 9.5      | 9.3      | −1.8 ± 0.1 | 1.4 ± 0.4         |
| GOME-2A       | Jan 2007 – Jun 2013 | 9.2      | 8.6      | −3.0 ± 0.2 | −0.3 ± 2.1        |
|               | Jul 2013 – Dec 2017 | 11.1     | 10.2     | 4.6 ± 0.2 | 1.5 ± 2.8         |
| GOME-2B       | Jan 2013 – Jun 2013 | 10.3     | 9.6      | −3.3 ± 0.4 | −59 ± 41          |
|               | Jul 2013 – Dec 2017 | 11.2     | 10.0     | 5.3 ± 0.2 | 2.7 ± 3.9         |
| OMPS-NOAA     | Mar 2012 – Dec 2014 | 22.0     | 19.2     | −9.3 ± 0.5 | 60 ± 16           |
|               | Jan 2015 – Dec 2017 | 10.2     | 9.7      | −2.6 ± 0.2 | −8.8 ± 6.3        |
| OMPS-NASA     | Nov 2011 – Dec 2017 | 8.8      | 8.8      | −0.2 ± 0.1 | 0.4 ± 1.1         |
| SVUV/N7       | Oct 1978 – Jun 1990 | 11.7     | 11.3     | 3.2 ± 0.3 | −3.2 ± 1.2        |
| SVUV/N9       | Feb 1985 – Aug 1988 | 11.7     | 11.7     | 0.6 ± 0.5 | −3.4 ± 6.6        |
| SVUV/N11      | Apr 1983 – Feb 1998 | 10.7     | 10.1     | 3.4 ± 0.3 | 4.0 ± 3.5         |
| SVUV/N14      | Dec 1988 – Jul 1993 | 11.1     | 10.9     | 1.8 ± 0.4 | 3.4 ± 3.2         |
|               | Nov 1997 – Mar 2001 | 11.7     | 11.5     | 1.8 ± 0.5 | 11 ± 7            |
| SVUV/N16      | Feb 1995 – Apr 2000 | 10.9     | 10.6     | 2.3 ± 0.3 | −2.1 ± 2.6        |
|               | Aug 2004 – Sep 2006 | 10.7     | 10.7     | −0.8 ± 0.5 | −10 ± 11          |
| SVUV/N17      | Oct 2000 – Jun 2007 | 10.8     | 10.7     | 1.4 ± 0.2 | −2.5 ± 1.5        |
|               | Jan 2012 – Jun 2014 | 11.0     | 11.0     | −0.1 ± 0.4 | 7.7 ± 6.9         |
| SVUV/N18      | Jul 2002 – Jan 2013 | 10.3     | 10.3     | 0.4 ± 0.2 | −1.4 ± 0.8        |
| SVUV/N19      | Jun 2005 – Dec 2012 | 10.1     | 10.0     | 1.1 ± 0.2 | 0.6 ± 1.1         |
|               | Feb 2009 – Jun 2016 | 10.4     | 10.3     | 0.8 ± 0.2 | 1.9 ± 1.2         |
| Merged TCO    | Nov 1978 – Dec 2017 | 8.6      | 8.4      |          |                   |

Errors of offset and drift are shown by twice the standard error.

Table 2. Simple linear regression (SLR) coefficients for differences between the satellite L2 overpass TCO (L2SAT) and ground-based TCO (GD), SAT–GD_{diff} = L2SAT − GD, as a function of time for the individual TCO datasets. The offset (intercept) and trend (slope) coefficients are determined from daily station-averaged SAT–GD_{diff} data. RMSD1 is for the uncorrected data, and RMSD2 is for the corrected data. Only overpass data collected when SZA < 84° and cloud cover < 80 % are used.
Fig. 1. Differences between the satellite L2 overpass total column ozone (TCO) (L2SAT) and ground-based TCO (GD) from a network of 71 Brewer and Dobson spectrophotometer stations, SAT–GDdiff = L2SAT − GD, as a function of time. Simple linear regression (SLR) lines in red that correspond to Table 2 are drawn. (a) TOMS/N7 (TN7), (b) TOMS/EP (TEP), (c) GOME (GOM), (d) TOGOMI (TGM), (e) SCIAMACHY (SCI), (f) TOSOMI (TSM), (g) OMI-TOMS (OMI), (h) OMI-DOAS (DOA), (i) GOME-2A (G2A), (j) GOME-2B (G2B), (k) OMPS-NOAA (ONO), (l) OMPS-NASA (ONA), (m) SBUV/N7 (S7), (n) SBUV/N9 (S9), (o) SBUV/N11 (S11), (p) SBUV/N14 (S14), (q) SBUV/N16 (S16), (r) SBUV/N17 (S17), (s) SBUV/N18 (S18), and (t) SBUV/N19 (S19). A 31-day running mean is applied.
changes on 1 July 2013, which coincided with the algorithm change from GDP 4.6 to GDP 4.7 (Figs. 1i, j) when the difference increases suddenly by ~ 8 DU. The seasonality associated with the data processed by DOAS-type algorithms can, for the most part, be attributed to the fact that the algorithms do not account for latitudinal and temporal changes in the effective temperature (Koukouli et al. 2016). In addition, data processed with DOAS-type algorithms acquired at high SZA show poor agreement with GD, apparently as a result of seasonal and geometrical problems (i.e., measurements at high SZA often located at high latitudes).

OMPS-NASA shows a remarkable agreement with GD, with a mean bias of around zero and a slightly upward trend of 0.4 % per decade (Fig. 1l). In contrast, during the early stage of satellite deployment in 2012–2014, OMPS-NOAA measurements are extremely degraded (Flynn et al. 2014), and SAT–GD$_{\text{diff}}$ shows an abnormally high slope. After 2015, however, the OMPS-NOAA SAT–GD$_{\text{diff}}$ exhibits stable behavior (Fig. 1k). Thus, we include only OMPS-NOAA from after 2015 in the merged TCO dataset.

The mean offsets of SAT–GD$_{\text{diff}}$ for datasets acquired with SBUV-type instruments are generally within 1–3 %. For daily data without smoothing (not shown), fewer good matches are available for pure nadir observation instruments than those for a two-dimensional CCD of OMI and a mapper of OMPS, and thus the variability in the SBUV anomalies is large. Because of poor sampling and low-quality SBUV instruments in the 1990s and early 2000s, SBUV/N9, /N11, and /N14 measurements exhibit large biases and significant drift due to large deviations of the equator crossing time (ECT) from an initial near-equator ECT (Kramarova et al. 2013; Frith et al. 2014). In early series SBUV datasets (SBUV/N7, /N9, /N11, and /N14; Figs. 1m-p), the peak-to-peak amplitude of 31-day-running mean fluctuations is within ±10 DU, whereas it is within ±5 DU in datasets acquired later (SBUV/N16, /N17, /N18, and /N19; Figs. 1q–t).

The latitudinal distributions of the mean relative SAT–GD$_{\text{diff}}$ i.e., (L2SAT − GD)/L2SAT, in the 20 datasets are shown in Fig. 2. Mean biases of satellite TCO relative to GD at each station range from −2 % to 5 %. The biases are small and close to zero at higher latitudes in NH and positive near the equator and in SH. However, the SH positive biases may be driven by the negative bias stations in NH mid-latitudes. The spread (variance) ranges from 3 % to 5 % at most of the stations. From the figure, no significant change is observed in biases in the GD records in the early part of the satellite records compared with the later part.

Ozone trends would be best diagnosed from the temporal evolution of the global TCO average. Figure 3 shows the time series of uncorrected satellite-measured TCO averaged over 60°S–60°N for each of the 20 datasets and the corresponding TCO data corrected by SLR as a function of time. Both the uncorrected and corrected time series show a general decline in ozone levels over time in the 1980s and early 1990s. The most visible effect of the trend correction is to reduce the trend over the period of 1978–1995.

Since the ground network was poorly calibrated in that period than in the post 2000 period, this might be due to uncertainty in the ground data. In this figure, there are a local maximum in 1988–1991 and a minimum in 1993. The low TCO values during 1992–1994, which are reduced by 5–10 DU compared with the 1990 level, are attributable to the Mount Pinatubo eruption in June 1991 (Fioletov et al. 2002). However, the TCO values return to the late 1980s level in 1998–1999, and no long-term drift is apparent after 2000.

The TOMS/N7 L2SAT overestimates GD by 2–4 DU during 1979–1993 (i.e., differences are positive; Fig. 1a). As a result, the corrected TCO magnitudes are reduced to 280–293 DU. Conversely, the TOMS/EP L2SAT underestimates GD during 1996–1999 (i.e., differences are negative; Fig. 1b). Thus, the corrected TCO values are higher than the uncorrected values. However, there are large differences between the corrected TOMS/EP dataset and other corrected data-sets which include data for 1996–1999 (i.e., GOME, TOGOMI, and SBUV/N9, /N11, and /N14; Figs. 1m-p), the peak-to-peak amplitude of 31-day-running mean fluctuations is within ±10 DU, whereas it is within ±5 DU in datasets acquired later (SBUV/N16, /N17, /N18, and /N19; Figs. 1q–t).

The satellite datasets for the 2000s agree well, within 5 DU, with one another, and the corrected datasets agree within 4 DU. In the 2010s, NASA-satellite TCO data (OMI-TOMS, OMI-DOAS, OMPS-NOAA, and OMPS-NASA) and SBUV-series data agree well with one another. However, GOME-2A and -2B TCO data are ~ 5 DU higher than the NASA-satellite and SBUV-series data (except for the OMPS-NOAA data from before 2015). The OMPS-NASA TCO dataset shows remarkable stability with respect to the ground-based network; the corrected globally-averaged OMPS-NASA TCO data are almost the same as those of the uncorrected data (Fig. 1l). Remarkably,
Fig. 2. Relative SAT–GD_{diff} for 20 individual TCO datasets as a function of the latitude of each ground-based station without bias correction. The averaged biases are denoted by red filled circles and standard deviations denoted by error bars. The datasets are the same as those shown in Fig. 1.
intersatellite differences between the NASA satellites and GOME-2 series after around 2013 significantly decrease in magnitude from ~5 DU for uncorrected TCOs to 2–3 DU for corrected TCOs, except for OMPS-NOAA.

In summary, the systematic mean biases and drift of TCO in the 20 satellite retrieval datasets as a function of time are removed by SLR fitting of the coincident satellite overpass and Dobson and Brewer GD data. In some datasets, measurements are extremely degraded during certain periods. In that case, linear regression analysis is applied separately to different parts of the satellite measurement period. In all the datasets, satellite overpass data agree with GD within ±2–3 % when the significantly degraded TOMS/EP measurements during 2000–2003 and OMPS-NOAA measurements during 2012–2014 are excluded. Abrupt jumps of ~8 DU in GOME-2A and -2B coincident with a change in the retrieval algorithm. The RMSD of the merged TCO dataset, 8.6 DU, is reduced to 8.4 DU by bias correction with respect to time.

5. Bias correction by multiple linear regression (MLR)

Biases of all the available satellite TCO datasets and the multi-sensor merged TCO dataset relative to the selected GDs are corrected as a function of time, SZA, and \( T_{\text{eff}} \) by MLR. First, number frequency distributions of satellite overpass measurements are presented for each dataset and then relative SAT–GD diff and RMSD are investigated as a function of SZA, cloud amount, and \( T_{\text{eff}} \). To assess the validity of this method for correcting biases in satellite TCO retrievals, we present the corrected TCO anomalies relative to uncorrected TCO values on a global map and examine the time series of globally averaged TCO.

In relation to SZA, frequency distributions of satellite overpass measurements (Fig. 4a) are unimodal or bimodal with peaks at SZA of 30–40° or 60° or both. The distributions of number frequencies and relative differences are heavily influenced by the fact that most of the ground stations are located in NH mid-latitudes.
Because the overpass SZA varies according to the season and latitude of the observations, the distribution largely reflects the geophysical location of the ground stations, which are mostly unevenly distributed in NH extratropics. Overall, the relative SAT–GD$_{\text{diff}}$ (Fig. 4b) and RMSD (Fig. 4c) are small when SZA is 20–60°. However, there are significant differences among the datasets when SZA is high, especially for SZA > 80°, because errors in the GD are likely due to scattered light.

The frequency distributions in relation to cloud amount (Fig. 4d) show cloud-free percentages of just 4–5 % for satellite overpass measurements retrieved by DOAS-type algorithms and 35–40 % for those by TOMS-type algorithms, except for OMPS-NOAA. The number frequencies of retrievals generally decrease as the cloud amount increases because overpass differences in this study use only GD in direct sun to ensure accuracy. Overall, the relative SAT–GD$_{\text{diff}}$ in most of the datasets shows little dependency on cloud amount (Fig. 4e). The SAT–GD$_{\text{diff}}$ in TOMS/N7 and GOME-2B (OMI-TOMS and OMPS-NOAA) shows positive (negative) biases, and biases of SCIAMACHY, TOSOMI, GOME-2A, and OMPS-NASA are near zero within ±1 %. The RMSD increases as increasing the cloud amount (Fig. 4f). Because clouds can substantially reduce the accuracy of satellite TCO products (Antón and Loyola 2011), these results suggest...
that cloud amount may be suitable for the consideration of TCO error rather than for the TCO itself.

The frequency distributions of satellite overpass measurements in relation to $T_{\text{eff}}$ (Fig. 4g) show that the L2 retrievals of $T_{\text{eff}}$ for GOME, SCIAMACHY, TOGOMI, GOME-2A and GOME-2B peak at ~ 230 K whereas the modal of JRA-55-derived $T_{\text{eff}}$ peaks at ~ 220 K. Van Roozendael et al. (2012) indicated that the DOAS may overestimate $T_{\text{eff}}$ and that a difference of 10 K in $T_{\text{eff}}$ results in a 3 % change in TCO. The $T_{\text{eff}}$ dependency of SAT–GD$_{\text{diff}}$ and its RMSD (Figs. 4h, i) show large spreads among the datasets at low temperatures because of the small sample sizes and large atmospheric variability in winter months. The relative SAT–GD$_{\text{diff}}$ for GOME, SCIAMACHY, TOGOMI, and TOSOMI shows larger positive biases at lower $T_{\text{eff}}$, whereas that for TOMS/EP, OMI-TOMS, and OMPS-NOAA shows large negative biases at lower $T_{\text{eff}}$. The OMI-DOAS, OMPS-NASA, and SBUV ensemble mean biases are small, within ±1 %, when $T_{\text{eff}}$ is higher than 220 K.

MLR coefficients of SAT–GD$_{\text{diff}}$ as a function of time, SZA, and $T_{\text{eff}}$ are presented in Table 3 for the 20 datasets, along with their RMSDs. The RMSDs of all the datasets are ~ 10 DU (3 %), and bias correction by MLR reduces the RMSD by 1–8 %, compared with the RMSDs of the uncorrected data (see RMSD1 in Table 2); GOME-2A and -2B show the largest RMSD reduction. The RMSD of uncorrected merged datasets, 8.6 DU, is reduced to 8.4 DU after correction by MLR.

Figure 5 shows latitude time cross-sections of differences between corrected TCO by MLR and uncorrected TCO for the 20 datasets (panels 5a–t) and the merged dataset (panel 5u). Overall, the correction differences show latitudinal dependency and distinct seasonal variation. The positive TCO correction in TOMS/EP, OMI-TOMS, OMI-DOAS, and OMPS-NOAA and negative TCO correction in GOME, TOGOMI, and GOME-2B at almost whole latitudes are basically attributed to negative and positive SAT–GD$_{\text{diff}}$, as shown in Fig. 1. Positive TCO correction at only higher latitudes is distinct in TOMS/N7. TCO correction is relatively small in SBUV datasets, due to small magnitude of SAT–GD$_{\text{diff}}$, as shown in Figs. 1 and 2. These general characteristics in individual data-

### Table 3. Multiple linear regression (MLR) coefficients for time, solar zenith angle (SZA), and effective ozone temperature ($T_{\text{eff}}$). RMSD3 is for SAT–GD$_{\text{diff}}$ corrected by MLR.

| Name           | RMSD3 DU | const DU | Time DU decade$^{-1}$ | SZA DU (10 deg)$^{-1}$ | $T_{\text{eff}}$ DU K$^{-1}$ |
|----------------|----------|----------|-----------------------|------------------------|-----------------------------|
| TOMS/N7        | 9.8      | 85.8     | -0.10                 | -0.56                  | 0.40                        |
| TOMS/EP         | 9.1      | 83.7     | -0.21                 | -0.23                  | 0.37                        |
| GOME           | 10.7     | 23.1     | -0.02                 | 0.26                   | -0.10                       |
| SCIAMACHY      | 8.5      | 11.3     | -0.02                 | 0.49                   | -0.06                       |
| TOGOMI         | 10.4     | 10.8     | -0.13                 | 0.49                   | -0.05                       |
| TOSOMI         | 8.8      | 1.8      | -0.15                 | 2.10                   | -0.05                       |
| OMI-TOMS       | 7.9      | 80.0     | 0.19                  | 0.44                   | 0.32                        |
| OMI-DOAS       | 9.0      | 95.4     | 0.14                  | 1.26                   | 0.39                        |
| GOME-2A        | 8.8      | 121      | 0.87                  | 2.15                   | 0.44                        |
| GOME-2B        | 9.4      | 132      | 0.73                  | 2.89                   | 0.48                        |
| OMPS-NOAA      | 9.6      | 76.3     | -0.61                 | 0.91                   | 0.36                        |
| OMPS-NASA      | 8.5      | 103      | 0.11                  | 1.06                   | 0.43                        |
| SBUV/N7        | 11.2     | 103      | -0.28                 | 0.83                   | 0.44                        |
| SBUV/N9        | 10.7     | 55.4     | 0.40                  | -0.17                  | 0.27                        |
| SBUV/N11       | 11.1     | 71.2     | 0.10                  | 0.37                   | 0.32                        |
| SBUV/N14       | 10.7     | 80.8     | -0.45                 | 1.02                   | 0.34                        |
| SBUV/N16       | 10.7     | 75.7     | -0.22                 | 0.46                   | 0.34                        |
| SBUV/N17       | 10.1     | 80.9     | -0.19                 | 1.07                   | 0.35                        |
| SBUV/N18       | 9.9      | 67.2     | 0.05                  | 0.56                   | 0.29                        |
| SBUV/N19       | 10.2     | 90.7     | 0.12                  | 0.79                   | 0.38                        |
| Merged TCO     | 8.4      |          |                       |                       |                             |

Data period: Tdemos/EP, July 1996 to December 1999 (but the bias is not corrected in the merged TCO data); GOME-2A, January 2007 to June 2013; GOME-2B, July 2013 to December 2017; and OMPS-NOAA, January 2015 to December 2017.
Fig. 5. Latitude time cross sections of differences between monthly mean corrected and uncorrected satellite TCOs (DU) in individual datasets. (a) TOMS/N7, (b) TOMS/EP, (c) GOME, (d) TOGOMI, (e) SCIAMACHY, (f) TOSOMI, (g) OMI-TOMS, (h) OMI-DOAS, (i) GOME-2A, (j) GOME-2B, (k) OMPS-NOAA, (l) OMPS-NASA, (m) SBUV/N7, (n) SBUV/N9, (o) SBUV/N11, (p) SBUV/N14, (q) SBUV/N16, (r) SBUV/N17, (s) SBUV/N18, and (t) SBUV/N19. (u) Merged TCO dataset during 1978–2017 (MRG).
Large gaps in SH after 2003 in the GOME and TOGOMI datasets are due to an ERS-2 recording failure (Koukouli et al. 2012). In the datasets processed by DOAS-type algorithms, TCO retrievals are not available when the Earth’s surface is covered by snow or ice, because the algorithms cannot distinguish between the snow- or ice-covered ground and cloud (Spurr et al. 2012). The effective light path is enhanced by multiple reflections in clouds, and when snow and ice is present, the highly reflecting surface produces an elongated “ghost” column; in both cases, TCO is overestimated (Loyola et al. 2011). However, in GOME-2A and -2B, after the retrieval algorithm was changed in mid-October 2016, the monthly coverage over 60–90°S increases to ~ 100 %, from ~ 40 % with the former algorithm.

It is interesting to examine how the spatial distribution and temporal behavior of the merged data change as a result of bias correction by MLR. For example, Figs. 6a–c show longitude-latitude sections of monthly mean corrected minus uncorrected TCO differences, RMSD, and $T_{\text{eff}}$ in October 2008. In general, TCO in October is high over the east Asia at higher latitudes and over the Southern Ocean, whereas it is low over 30°N–30°S region and over Antarctica. A large part of the merged TCO data in 2008 is derived from the OMI-TOMS and OMI-DOAS datasets, both of which have negative biases. Thus, the corrected TCO results show increases in TCO in most parts of the world (Fig. 6a). Figure 6b shows that the RMSD is large at the equator and in Antarctica, and small in the Southern Ocean in this example.

Figure 6c shows a low $T_{\text{eff}}$ in lower latitudes and a high $T_{\text{eff}}$ in the Southern Ocean. In Table 3, coefficients of $T_{\text{eff}}$ for OMI-TOMS and OMI-DOAS are positive,
indicating that the bias correction in higher/lower $T_{eff}$ regions show decreases/increases in TCO in those regions. The key point is that this bias correction by MLR is not uniform globally because the correction in a global map depends on $T_{eff}$ and SZA, which are critical parameters affecting TCO retrievals (van der A et al. 2010).

The time series of bias correction by SLR and MLR at a mid-latitude station in NH is shown in Fig. 6d. Bias correction by SLR does not reveal any seasonal behavior because this correction uses only the mean offset and drift. In contrast, bias correction by MLR reveals clear seasonal behavior. The seasonal peak-to-peak amplitude changes from $\sim 3$ DU before 2013 and to $\sim 6$ DU after 2013. This amplitude change coincides with the algorithm change in GOME-2A and -2B. The GOME-2A and -2B satellite overpass time series (Figs. 1i, j, respectively) show negative biases against GD before 2013 and positive biases after 2013.

Figure 7 shows the time series of merged uncorrected and corrected TCOs by SLR and MLR averaged over 60°S–60°N, and the differences of corrected TCO relative to the uncorrected TCO; NH and SH averages are also shown. For comparison, the KNMI ozone multi-sensor reanalysis of van der A et al. (2010) is plotted. The reanalysis was constructed by data assimilation using reprocessed versions of 13 satellite corrected datasets for 1978–2008 against GD for biases as a function of time, SZA, VZA, and $T_{eff}$.

The behavior of the global, NH and SH averages of TCO corrected by SLR is similar to that corrected by MLR, and good agreement is also seen between both our corrected and merged TCO time series and the merged KNMI time series of van der A et al. (2010). This agreement indicates that the simplified method in which only the mean satellite bias and drift are corrected has some benefits as an assimilation data source in that it removes the discontinuity caused by satellite
de-biased data, and it is expected to reduce apparent biases among individual satellite measurements. From Fig. 7b, the bias correction by MLR is higher than that by SLR. This difference could be caused by $T_{eff}$ of Dobson and retrieved satellite ozone as well as SZA. By comparing these time series with those of the individual satellite datasets (Fig. 3), we find that both SLR and MLR fittings correct the positive biases during the TOMS/N7 observation period of 1978–1993 and the negative biases during the OMI period of 2004–2017.

In summary, bias correction by MLR is not uniform globally and shows distinct seasonal variation because the correction depends on time, SZA, and $T_{eff}$. The bias correction by MLR is higher than that by SLR. In addition, the time series after SLR correction for mean bias and drift shows good agreement with the TCO time series corrected by MLR ($R^2 = 0.998$). Therefore, the simplified method has some benefits as an assimilation data source in that it removes the discontinuity caused by satellite de-biased data, and it is expected to reduce apparent biases among individual satellite measurements.

### 6. Trend analysis

In this section, the TCO trend in the corrected and uncorrected merged TCO datasets is investigated. Apparent biases in individual satellite measurements are reduced in the MLR corrected merged dataset.

Figure 8 shows the seasonal march of the TCO trend in NH and SH mid-latitudes along with the latitudinal distribution of the TCO trend during two periods, 1979–1996 and 2000–2017. These two different trends correspond to the ozone decrease in the 1980s and 1990s and the slow ozone increase since the early 2000s (Petropavlovskikh et al. 2019).

Fig. 8. Seasonal evolution of the TCO trend over (a) 35–60°N and (b) 35–60°S during 1979–1996 (blue) and 2000–2017 (red) in the merged dataset corrected by MLR. Units are DU yr$^{-1}$. The trend without correction is shown in black. Error bars (long-short dashed lines) denote twice the standard error of the mean trend with (without) bias correction. Latitudinal distribution of the TCO trend during (c) 1979–1996 (blue) and (d) 2000–2017 (red) with bias correction. The trend without correction is shown in black.
ozone decline is less seasonally dependent. Overall, TCO trend shows a statistically significant decline of 10–20 DU per decade during 1979–1996 (Figs. 8a, b). During 2000–2017, the trend is near zero in all seasons in both NH and SH mid-latitudes, except for an insignificant positive trend in January over 35–60°N and a significant positive trend in April and May over 35–60°S. In both hemispheres during 1979–1996, the difference between the corrected and uncorrected TCO trends shows seasonal dependence, being positive in summer months. This influence may be attributed to \( T_{\text{eff}} \) and SZA dependency for MLR. It is notable that bias correction by SLR loses seasonality for differences between corrected and uncorrected TCOs (not shown).

The TCO trend is near zero over the equator and substantially declines outside the 30°S–30°N zone during 1979–1996 (Fig. 8c). The cumulative ozone loss over the mid-latitude during that period is ~ 5 %, in agreement with results of Fioletov et al. (2002) and World Meteorological Organization (2015). In this period, the corrected TCO data show smaller negative trends compared with the uncorrected data. During 2000–2017 (Fig. 8d), the trend is near zero over 20–30°S and 10–60°N. Meanwhile, it is significantly positive over 10°S–10°N and insignificantly positive over 40–60°S. In this period, the corrected data show a large positive trend over 10°S–30°N, compared with the uncorrected data. Overall, our trend correction by MLR is higher than that by SLR especially at lower latitudes because bias correction increases in TCO at low \( T_{\text{eff}} \).

In summary, the difference between MLR corrected and uncorrected TCO trends shows clear seasonal and latitudinal dependency while such seasonal and latitudinal dependence of bias correction by SLR against time is lost. The differences between MLR and SLR corrected TCO trends are generally positive at lower latitudes where bias correction by MLR increases in TCO due to low \( T_{\text{eff}} \).

7. Concluding remarks

In this study, we have constructed the merged dataset using 20 available satellite TCO datasets covering the 40 years from 1978 to 2017. The individual datasets and the merged dataset are corrected against selected GD by two methods: SLR against time and MLR against time, SZA, and \( T_{\text{eff}} \).

In the 20 datasets, SAT–GD shows little dependence of SZA when SZA < 60°. However, it shows large deviations when SZA > 80° with anomalies > 5 % because GD errors are likely due to scattered light at large SZA. The dependency of the cloud amount is small in most of the datasets and the (TCO) RMSD for the cloud amount increases as increasing the cloud amount. Although clouds can significantly reduce the accuracy of satellite TCO retrievals (Antón and Loyola 2011), this means that cloud amount may not be a suitable parameter for TCO bias correction and that it may be taken into consideration in TCO error correction. SAT–GD shows a large spread at low \( T_{\text{eff}} \), and this result is attributable to small sample sizes and probably also to large atmospheric variability in winter months.

Because SZA and \( T_{\text{eff}} \) dependencies are geophysically linked (Koukouli et al. 2016), these two effects are important for validation studies. van der A et al. (2010) also indicated that SZA and \( T_{\text{eff}} \) have clear seasonal components and that an explicit seasonal or latitudinal dependency almost disappears when their corrections were applied to the satellite observations. In our analysis of SAT–GD, all datasets agree within ±2–3 % (< 10 DU), except for some significantly degraded measurements in TOMS-EP during 2000–2003 and OMPS-NOAA during 2012–2014. The systematic drifts are within ±1 % (< 4 DU) per decade, except in degraded TOMS/EP measurements, and in OMPS-NOAA and GOME-2B measurements during the early stage of satellite deployment. Later, OMPS-NOAA and GOME-2B retrievals exhibit stable behaviors.

The merged datasets are created by averaging all data located within a grid cell after multiplication with weights inversely proportional to TCO errors. The behavior of the merged TCO after bias correction by SLR is similar to that after correction by MLR. Both corrected merged TCO datasets agree well with the KNMI merged TCO reported by van der A et al. (2010). Therefore, correction for mean bias and drift also improves the merged TCO datasets. The RMSDs in the corrected datasets by SLR and MLR are reduced from 8.6 DU to 8.4 DU. This is an additional benefit for data correction and time-series homogenization. A drawback of bias correction by SLR is that bias correction of seasonal and latitudinal dependence is not detected, whereas bias correction by MLR shows clear seasonal and latitudinal dependence (Fig. 6).

For the trend analysis, bias correction by MLR shows clear seasonal and latitudinal dependence while such seasonal and latitudinal dependence of bias correction by SLR against time is lost. The differences between MLR corrected TCO and uncorrected TCO trends are generally positive at lower latitudes where bias correction increases in TCO due to low \( T_{\text{eff}} \). A further investigation is needed for this positive TCO
correction in the low latitude.

Data availability. The homogenized monthly mean merged TCO datasets are available from the authors upon request via e-mail (biascorrection_l2sat_tatol ozone AT mri-jma.go.jp).

Acknowledgments

This research was supported by the Environment Research and Technology Development Fund (S-12) of the Environmental Restoration and Conservation Agency of Japan and by Grant-in-Aid (JP18K03748) for Science Research (KAKENHI) from the Ministry of Education, Culture, Sports, Science, and Technology of Japan. The authors give many thanks to Atsuya Kinoshita, Itaru Uesato, Mikio Ueno, Nozomu Ohkawara, Makoto Inoue, Toshihiko Hirooka, and especially Kazuyuki Miyazaki. Without his constructive and scientific advices to claim JMA’s ozone products, nothing about our ozone homogenization project would be accomplished at all.

Appendix

Table A1. List of Brewer stations. Station’s latitude, a time period of operation (from Year1 to Year2), and a single or double Brewer-data type are shown.

| ID | Station Name         | Latitude | Year1 | Year2 | Single or Double |
|----|----------------------|----------|-------|-------|-----------------|
| 002| Tamanrasset          | 22.8°N   | 2011  | 2017  | D               |
| 018| Alert                | 82.5°N   | 1989  | 2016  | S               |
| 021| Edmonton             | 53.6°N   | 1984  | 2016  | S → D           |
| 024| Resolute             | 74.7°N   | 1988  | 2016  | S → D           |
| 030| Minamitorishima      | 24.3°N   | 1997  | 2016  | S → D           |
| 050| Potsdam              | 52.2°N   | 1987  | 2003  | S               |
| 053| Uccle                | 50.8°N   | 1983  | 2017  | S → D           |
| 065| Toronto              | 43.8°N   | 1984  | 2014  | S → D           |
| 076| Goose Bay            | 53.3°N   | 1985  | 2016  | S → D           |
| 077| Churchill            | 58.8°N   | 1986  | 2016  | S → D           |
| 096| Hradec Králové       | 50.2°N   | 1994  | 2017  | S → D           |
| 099| Hohenpeissenberg     | 47.8°N   | 1984  | 2017  | S               |
| 111| South Pole           | 90.0°S   | 2008  | 2016  | D               |
| 165| Oslo                 | 59.9°N   | 1990  | 2017  | S               |
| 174| Lindenberg           | 52.2°N   | 1992  | 2014  | S               |
| 213| El Arenosillo        | 37.1°N   | 2000  | 2012  | D               |
| 233| Marambio             | 64.2°S   | 2010  | 2017  | D               |
| 261| Thessaloniki         | 40.5°N   | 1982  | 2017  | S               |
| 262| Sodankylä           | 67.4°N   | 1988  | 2010  | S               |
| 267| Sondrestrom          | 67.0°N   | 1995  | 2017  | S               |
| 279| Norrköping           | 58.6°N   | 1988  | 2017  | S               |
| 284| Vindeln              | 64.2°N   | 1996  | 2017  | S               |
| 290| Saturna Island       | 48.8°N   | 1996  | 2016  | D               |
| 308| Madrid               | 40.5°N   | 1991  | 2017  | D               |
| 315| Eureka               | 80.1°N   | 2001  | 2016  | S               |
| 316| De Bilt              | 52.1°N   | 1994  | 2017  | D               |
| 318| Valentia            | 51.9°N   | 1993  | 2017  | S → D           |
| 331| Poprad-Ganovec       | 49.0°N   | 1993  | 2017  | S               |
| 346| Murcia               | 38.0°N   | 1995  | 2017  | S               |
| 352| Manchester           | 53.5°N   | 2000  | 2017  | D               |
| 353| Reading              | 51.4°N   | 2002  | 2017  | S               |
| 411| Zaragoza            | 41.6°N   | 2000  | 2017  | S               |
| 435| Paramaribo           | 5.8°N    | 1999  | 2017  | D               |
| 479| Aosta               | 45.7°N   | 2007  | 2017  | S               |
Table A2. List of Dobson stations. Station’s latitude and a time period of operation (from Year1 to Year2).

| ID  | Station Name              | Latitude | Year1 | Year2 |
|-----|--------------------------|---------|-------|-------|
| 002 | Tamanrasset              | 22.8°N  | 1957  | 2017  |
| 010 | New Delhi                | 28.7°N  | 1957  | 2015  |
| 012 | Sapporo                  | 43.1°N  | 1958  | 2017  |
| 014 | Tateno                   | 36.1°N  | 1957  | 2017  |
| 019 | Bismarck                 | 46.8°N  | 1957  | 2017  |
| 020 | Caribou                  | 46.9°N  | 1958  | 2017  |
| 027 | Brisbane                 | 27.4°S  | 1986  | 2017  |
| 029 | Macquarie Island         | 54.5°S  | 1990  | 2017  |
| 040 | Haute Provence           | 43.9°N  | 1959  | 2017  |
| 043 | Lerwick                  | 60.1°N  | 1952  | 2017  |
| 050 | Potsdam                  | 52.2°N  | 1964  | 2003  |
| 051 | Reykjavik                | 64.1°N  | 1952  | 2017  |
| 053 | Ucle                     | 50.8°N  | 1952  | 2009  |
| 057 | Halley                   | 75.5°S  | 1957  | 2017  |
| 065 | Toronto                  | 43.8°N  | 1960  | 2001  |
| 067 | Boulder ESRL             | 40.0°N  | 1964  | 2017  |
| 068 | Belsk                    | 51.8°N  | 1963  | 2017  |
| 084 | Darwin                   | 12.4°S  | 1966  | 2017  |
| 096 | Hradec Králové           | 50.2°N  | 1961  | 2017  |
| 099 | Hohenpeissenberg         | 47.8°N  | 1967  | 2017  |
| 101 | Syowa                    | 69.0°S  | 1961  | 2017  |
| 105 | Fairbanks                | 64.8°N  | 1952  | 2017  |
| 106 | Nashville                | 36.3°N  | 1962  | 2017  |
| 107 | Wallops                  | 37.9°N  | 1967  | 2017  |
| 111 | South Pole               | 90.0°S  | 1961  | 2017  |
| 159 | Perth                    | 31.9°S  | 1969  | 2016  |
| 190 | Naha                     | 26.2°N  | 1974  | 2017  |
| 191 | Samoa                    | 14.3°S  | 1975  | 2017  |
| 199 | Barrow                   | 71.3°N  | 1973  | 2017  |
| 208 | Xianghe                  | 40.0°N  | 1979  | 2017  |
| 213 | El Arenosillo            | 37.1°N  | 1976  | 2012  |
| 214 | Singapore                | 1.3°N   | 1979  | 1995  |
| 253 | Melbourne                | 37.8°S  | 1983  | 2017  |
| 268 | Arrival Heights          | 77.8°S  | 1988  | 2016  |
| 284 | Vindeln                  | 64.2°N  | 1991  | 2017  |
| 340 | Springbok                | 29.7°S  | 1995  | 2016  |
| 342 | Comodoro Rivadavia       | 45.8°S  | 1995  | 2016  |
| 409 | Hurghada                 | 27.3°N  | 2001  | 2017  |
| 498 | Kyiv-Goloseyev           | 50.4°N  | 2010  | 2017  |
References

Antón, M., and D. Loyola, 2011: Influence of cloud properties on satellite total ozone observations. *J. Geophys. Res.*, **116**, D03208, doi:10.1029/2010JD014780.

Antón, M., M. E. Koukouli, M. Kroon, R. D. McPeters, G. J. Labow, D. Balis, and A. Serrano, 2010: Global validation of empirically corrected EP-Total Ozone Mapping Spectrometer (TOMS) total ozone columns using Brewer and Dobson ground-based measurements. *J. Geophys. Res.*, **115**, D19305, doi:10.1029/2010JD014178.

Bhattia, P. K., 2002: *OMI algorithm theoretical basis document Volume II OMI ozone products*. ATBD-OMI-02, Version 2.0, NASA, USA, 91 pp.

Bhattia, P. K., 2005: *OMI/Aura Ozone (O3) Total Column I-Orbit L2 Swath 13x24 km V003*. Goddard Earth Sciences Data and Information Services Center (GES DISC), NASA, USA, doi:10.5067/Aura/OMI/DATA2024. [Available at https://disc.gsfc.nasa.gov/datasets/OMTO3_003/]

Bhattia, P. K., R. D. McPeters, L. E. Flynn, S. Taylor, N. A. Kramarova, S. Frith, B. Fisher, and M. DeLand, 2013: Solar Backscatter UV (SBUV) total ozone and profile algorithm. *Atmos. Meas. Tech.*, **6**, 2533–2548.

Bodeker, G. E., J. C. Scott, K. Kreher, and R. L. McKenzie, 2001: Global ozone trends in potential vorticity coordinates using TOMS and GOME intercompared against the Dobson network: 1978–1998. *J. Geophys. Res.*, **106**, 23029–23042.

Damadeo, R. P., J. M. Zawodny, and L. W. Thomason, 2014: Reevaluation of stratospheric ozone trends from SAGE II data using a simultaneous temporal and spatial analysis. *Atmos. Chem. Phys.*, **14**, 13455–13470.

Deutsches Zentrum für Luft und Raumfahrt, 2012: *Product user manual for GOME total columns of ozone, NO3, tropospheric NO2, BrO, SO2, H2O, HCIO, OCIO, and Cloud Properties*. DLR/GOME/PUM/01, Germany, 47 pp. [Available at https://atmos.eoc.dlr.de/]

Dragani, R., 2011: On the quality of the ERA-Interim ozone reanalyses: Comparisons with satellite data. *Quart. J. Roy. Meteor. Soc.*, **137**, 1312–1326.

Ebojie, F., C. von Savigny, A. Ladstätter-Weißenmayer, A. Rozanov, M. Weber, K.-U. Eichmann, S. Bötel, N. Rahpoe, H. Bovensmann, and J. P. Burrows, 2014: Tropospheric column amount of ozone retrieved from SCIAMACHY limb-nadir-matching observations. *Atmos. Meas. Tech.*, **7**, 2073–2096.

Eskes, H. J., R. J. van der A, A. E. J. Brinksma, J. P. Vreekind, 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., G. E. Bodeker, A. J. Miller, R. D. McPeters, and R. Stolarski, 2002: Global and zonal total ozone variations estimated from ground-based and satellite measurements: 1964–2000. *J. Geophys. Res.*, **107**, 4647, doi:10.1029/2001JD001350.

Fioletov, V. E., G. Labow, R. Evans, E. W. Hare, U. Köhler, C. T. McElroy, K. Miyagawa, A. Redondas, V. Savastiouk, A. M. Shalayamsky, J. Staehelin, K. Vanicek, and M. Weber, 2008: Performance of the ground-based total ozone network assessed using satellite data. *J. Geophys. Res.*, **113**, D14313, doi:10.1029/2008JD009809.

Flynn, L., C. Long, X. Wu, R. Evans, C. T. Beck, I. Petrovlovskikh, G. McConville, W. Yu, Z. Zhang, J. Niu, E. Beach, Y. Hao, C. Pan, B. Sen, M. Novicki, S. Zhou, and C. Seftor, 2014: Performance of the Ozone Mapping and Profiler Suite (OMPS) products. *J. Geophys. Res.*, **119**, 6181–6195.

Fortuin, J. P. F., and H. Kelder, 1998: An ozone climatology base on ozoneonde and satellite measurements. *J. Geophys. Res.*, **103**, 31709–31734.

Frith, S. M., N. A. Kramarova, R. S. Stolarski, R. D. McPeters, P. K. Bhattia, and G. J. Labow, 2014: Recent changes in total column ozone based on the SBUV Version 8.6 Merged Ozone Data Set. *J. Geophys. Res.*, **119**, 9735–9751. [Available at https://acd-ext.gsfc.nasa.gov/anonymous/toms/sbuv/]

Goddin, R., 2014: *Joint Polar Satellite System (JPSS) OMPS N4DIR Total Column Ozone Algorithm Theoretical Basis Document (ATBD)*. 474-00029, Goddard Space Flight Center, USA. [Available at https://www.class.ngdc.noaa.gov/]

Jaross, G., 2017: *OMPS-NPP L2 NM Ozone (O3) Total Column swath orbital V2*. Goddard Earth Sciences Data and Information Services Center (GES DISC), NASA, USA. [Available at https://disc.gsfc.nasa.gov/datasets/OMPS_NPP_NMTO3_L2_2/]

Karppinen, T., A. Redondas, R. D. García, K. Lakkala, C. T. McElroy, and E. Kyrö, 2015: Compensating for the effects of stray light in single-monochromator Brewer spectrophotometer ozone retrieval. *Atmos.-Ocean*, **53**, 66–73.

Kobayashi, S., Y. Ota, Y. Harada, A. Ebita, M. Moriya, H. Onoda, K. Onogi, H. Kamahori, C. Kobayashi, H. Endo, K. Miyaoaka, and K. Takahashi, 2015: The JRA-55 Reanalysis: General specifications and basic characteristics. *J. Meteor. Soc. Japan*, **93**, 5–48.

Koukouli, M. E., C. Lerot, J. Granville, F. Goutail, J.-P. Pommereau, D. Balis, I. Zyrichidou, M. Van Roozendael, C. Lerot, and R. J. D. Spurr, 2012: Geophysical validation and long-term consistency between GOME-2/ MetOp-A total ozone column and measurements from the sensors GOME/ERS-2, SCIAMACHY/ENVISAT and OMI/Aura. *Atmos. Meas. Tech.*, **5**, 2169–2181.

Koukouli, M. E., C. Lerot, J. Granville, F. Goutail, J.-P. Pommereau, D. Balis, I. Zyrichidou, M. Van Roozendael, M. Coldewey-Egbers, D. Loyola, G. Labow, S. Frith, R. Spurr, and C. Zehner, 2015: Evaluating a new homogeneous total ozone climate data...
record from GOME/ERS-2, SCIAMACHY/Envisat, and GOME-2/MetOp-A. J. Geophys. Res., 120, 12296–12312.

Koukouli, M. A., S. M. Frith, K. Fagkos, D. Balis, M. Van Roozendael, M. A. F. Allart, and R. J. van der A, 2016: The impact of the ozone effective temperature on satellite validation using the Dobson spectrophotometer network. Atmos. Meas. Tech., 9, 2055–2065.

Kramarova, N. A., E. R. Nash, P. A. Newman, P. K. Bhartia, D. F. Rault, C. J. Seftor, P. Q. Xu, and G. J. Labow, 2014: Measuring the Antarctic ozone hole with the new Ozone Mapping and Profiler Suite (OMPS). Atmos. Chem. Phys., 14, 2353–2361.

Kroon, M., J. P. Veefkind, M. Sneep, R. D. McPeters, P. K. Bhartia, and P. F. Levelt, 2008: Comparing OMITOMS and OMI-DOAS total ozone column data. J. Geophys. Res., 113, D16S28, doi:10.1029/2007JD008798.

Kroon, M., J. F. de Haan, J. P. Veefkind, L. Froidevaux, R. Wang, R. Kivi, and J. J. Hakkarainen, 2011: Validation of operational ozone profiles from the Ozone Monitoring Instrument. J. Geophys. Res., 116, D18305, doi:10.1029/2010JD015100.

Lerot, C., M. Van Roozendael, J. van Geffen, J. van Gent, C. Fayt, R. Spurr, G. Lichtenberg, and A. von Bargen, 2009: Six years of total ozone column measurements from SCIAMACHY nadir observations. Atmos. Meas. Tech., 2, 87–98.

Lerot, C., M. Van Roozendael, J.-C. Lambert, J. Granville, J. van Gent, D. Loyola, and R. Spurr, 2010: The GODFIT algorithm: A direct fitting approach to improve the accuracy of total ozone measurements from GOME. Int. J. Remote Sens., 31, 543–550.

Lerot, C., M. Van Roozendael, R. Spurr, D. Loyola, M. Coldewey-Egbers, S. Kochenova, J. van Gent, M. Koukouli, D. Balis, J.-C. Lambert, J. Granville, and C. Zehner, 2014: Homogenized total ozone data records from the European sensors GOME/ERS-2, SCIAMACHY/Envisat, and GOME-2/MetOp-A. J. Geophys. Res., 119, 1639–1662.

Loyola, D. G., M. E. Koukouli, P. Valks, D. S. Balis, N. Hao, M. Van Roozendael, R. J. D. Spurr, W. Zimmer, S. Kiemle, C. Lerot, and J.-C. Lambert, 2011: The GOME-2 total column ozone product: Retrieval algorithm and ground-based validation. J. Geophys. Res., 116, D07302, doi:10.1029/2010JD014675.

McPeters, R. D., P. K. Bhartia, A. J. Krueger, J. R. Herman, B. M. Schlesinger, C. G. Wellemeyer, C. J. Seftoret, G. Jaross, S. L. Taylor, T. Swissler, O. Torres, G. Labow, W. Byerly, and R. P. Cebula, 1996: Nimbus-7 Total Ozone Mapping Spectrometer (TOMS) Data Products User’s Guide. NASA Reference Publication, USA, 67 pp.

McPeters, R. D., P. K. Bhartia, A. J. Krueger, J. R. Herman, C. G. Wellemeyer, C. J. Seftoret, G. Jaross, O. Torres, L. Moy, G. Labow, W. Byerly, S. L. Taylor, T. Swissler, and R. P. Cebula, 1998: Earth-Probe Total Ozone Mapping Spectrometer (TOMS) Data Products User’s Guide. NASA Reference Publication, USA, 60 pp.

McPeters, R., M. Kroon, G. Labow, E. Brinksma, D. Balis, I. Petropavlovskikh, J. P. Veefkind, P. K. Bhartia, and P. F. Levelt, 2008: Validation of the Aura Ozone Monitoring Instrument total column ozone product. J. Geophys. Res., 113, D15S14, doi:10.1029/2007JD008802.

McPeters, R. D., P. K. Bhartia, D. Haffner, G. J. Labow, and L. Flynn, 2013: The version 8.6 SBUV ozone data record: An overview. J. Geophys. Res., 118, 8032–8039.

Miller, A. J., R. M. Nagatani, L. E. Flynn, S. Kondragunta, E. Beach, R. Stolarski, R. D. McPeters, P. K. Bhartia, M. T. DeLand, C. H. Jackman, D. J. Wuebbles, K. O. Patten, and R. P. Cebula, 2002: A cohesive total ozone data set from the SBUV/2 satellite system. J. Geophys. Res., 107, 4701, doi:10.1029/2001JD000853.

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

OMI Team, 2012: Ozone Monitoring Instrument (OMI) data user’s guide. OMI-DUG-5.0, NASA, USA, 62 pp.

Ozone CCI, 2012: Product specification document (PSD). Ozone_cci_PSD_3.1, Germany, 41 pp. [Available at http://www.esa-ozone-cci.org/]

Petropavlovskikh, I., S. Godin-Beekmann, D. Hubert, R. Damadeo, B. Hassler, V. Sofieva, S. M. Frith, and K. Tourpali, 2019: SPARC/IO3C/GAW Report on long-term ozone trends and uncertainties in the stratosphere. SPARC Report No. 9, GAW Report No. 241, WCRP-17/2018, 78 pp.

Spurr, R., M. van Roozendael, D. Loyola, C. Lerot, J. van Geffen, J. van Gent, C. Fayt, J.-C. Lambert, W. Zimmer, A. Doicu, S. Otto, D. Balis, M. Koukouli, and C. Zehner, 2012: GDP 5.0 - Upgrade of the GOME Data Processor for improved total ozone columns, algorithm theoretical basis document. GDP5-ATBD-1, DLR, Germany, 71 pp.

Stolarski, R. S., and S. M. Frith, 2006: Search for evidence of trend slow-down in the long-term TOMS/SBUV total ozone data record: the importance of instrument drift uncertainty. Atmos. Chem. Phys., 6, 4057–4065.

Tropospheric Emission Monitoring Internet Service, 2005: Service report protocol monitoring. SR3/002, TEMIS, Netherland, 30 pp.

TOMS Science Team, 1996: TOMS/7L2: TOMS Nimbus-7 Ozone (O3) Total Column I-Orbit L2 Swath 50 km x
Van Roozendael, M., R. Spurr, D. Loyola, C. Lerot, D. Balis, J.-C. Lambert, W. Zimmer, J. van Gent, J. van Geffen, M. Koukouli, J. Granville, A. Doicu, C. Fayt, and C. Zehner, 2012: Sixteen years of GOME/ERS-2 total ozone data: The new direct-fitting GOME Data Processor (GDP) version 5—Algorithm description. J. Geophys. Res., 117, D03305, doi:10.1029/2011JD016471.

Van der A, R. J., and H. J. Eskes, 2006: Product specification document, TOSOMI. PSD3, TEMIS, Netherland, 13 pp. [Available at http://www.temis.nl/protocols/O3total.html/]

van der A, R. J., and H. J. Eskes, 2010: Multi sensor reanalysis of total ozone. Atmos. Chem. Phys., 10, 11277–11294.

van der A, R. J., M. A. F. Allaart, and H. J. Eskes, 2015: Extended and refined multi sensor reanalysis of total ozone for the period 1970–2012. Atmos. Meas. Tech., 8, 3021–3035.

Van Roozendael, M., D. Loyola, R. Spurr, D. Balis, J.-C. Lambert, Y. Livschitz, P. Valks, T. Ruppert, P. Kenter, C. Fayt, and C. Zehner, 2006: Ten years of GOME/ERS-2 total ozone data — The new GOME data processor (GDP) version 4: 1. Algorithm description. J. Geophys. Res., 111, D14311, doi:10.1029/2005JD006375.

Valks, P. J. M., J. F. De Haan, J. P. Veeckind, R. F. van Oss, and D. S. Balis, 2004: TOGOMI: An improved total ozone retrieval algorithm for GOME. XX Quadrennial Ozone Symposium. Zerefos, C. S. (ed.), University of Athens, 129–130. [Available at http://www.temis.nl/protocols/O3total.html/]

Veefkind, P., 2006: OMI/Aura Ozone (O3) DOAS Total Column 1-Orbit L2 Swath 13x24 km V003. Goddard Earth Sciences Data and Information Services Center (GES DISC), NASA, USA. [Available at https://disc.gsfc.nasa.gov and a dataset name is OMDOAO3_003.]

Wargan, K., G. Labow, S. Frith, S. Pawson., N. Livesey, and G. Partyka, 2017: Evaluation of the ozone fields in NASA’s MERRA-2 Reanalysis. J. Climate, 30, 2961–2988.

World Meteorological Organization, 2015: Scientific Assessment of Ozone Depletion: 2014. World Meteorological Organization, Global Ozone Research and Monitoring Project, Rep. 55, 416 pp.

World Meteorological Organization, 2019: Scientific Assessment of Ozone Depletion: 2018. World Meteorological Organization, Global Ozone Research and Monitoring Project, Rep. 58, 588 pp.

50 km V008. Goddard Earth Sciences Data and Information Services Center (GES DISC), NASA, USA. [Available at https://disc.gsfc.nasa.gov/datacollection/TOMSN7L2_008.html.]

TOMS Science Team, 1998: TOMSEPL2: TOMS Earth-Probe Ozone (O3) Total Column 1-Orbit L2 Swath 50 x 50 km V008. Goddard Earth Sciences Data and Information Services Center (GES DISC), NASA, USA. [Available at https://disc.gsfc.nasa.gov/datacollection/TOMSEPL2_008.html.]

Valks, P. J. M., J. F. De Haan, J. P. Veefkind, R. F. van Oss, and D. S. Balis, 2004: TOGOMI: An improved total ozone retrieval algorithm for GOME. XX Quadrennial Ozone Symposium. Zerefos, C. S. (ed.), University of Athens, 129–130. [Available at http://www.temis.nl/protocols/O3total.html.]

van der A, R. J., M. A. F. Allaart, and H. J. Eskes, 2010: Multi sensor reanalysis of total ozone. Atmos. Chem. Phys., 10, 11277–11294.

van der A, R. J., M. A. F. Allaart, and H. J. Eskes, 2015: Extended and refined multi sensor reanalysis of total ozone for the period 1970–2012. Atmos. Meas. Tech., 8, 3021–3035.

Van Roozendael, M., D. Loyola, R. Spurr, D. Balis, J.-C. Lambert, Y. Livschitz, P. Valks, T. Ruppert, P. Kenter, C. Fayt, and C. Zehner, 2006: Ten years of GOME/ERS-2 total ozone data — The new GOME data processor (GDP) version 4: 1. Algorithm description. J. Geophys. Res., 111, D14311, doi:10.1029/2005JD006375.

Van Roozendael, M., R. Spurr, D. Loyola, C. Lerot, D. Balis, J.-C. Lambert, W. Zimmer, J. van Gent, J. van Geffen, M. Koukouli, J. Granville, A. Doicu, C. Fayt, and C. Zehner, 2012: Sixteen years of GOME/ERS-2 total ozone data: The new direct-fitting GOME Data Processor (GDP) version 5—Algorithm description. J. Geophys. Res., 117, D03305, doi:10.1029/2011JD016471.

Veenkind, P., 2006: OMI/Aura Ozone (O3) DOAS Total Column 1-Orbit L2 Swath 13x24 km V003. Goddard Earth Sciences Data and Information Services Center (GES DISC), NASA, USA. [Available at https://disc.gsfc.nasa.gov and a dataset name is OMDOAO3_003.]

Veenkind, J. P., J. F. de Haan, E. J. Brinksma, M. Kroon, and P. F. Levelt, 2006: Total ozone from the Ozone Monitoring Instrument (OMI) using the DOAS technique. IEEE Trans. Geosci. Remote Sens., 44, 1239–1244.

World Meteorological Organization, 2015: Scientific Assessment of Ozone Depletion: 2014. World Meteorological Organization, Global Ozone Research and Monitoring Project, Rep. 55, 416 pp.

World Meteorological Organization, 2019: Scientific Assessment of Ozone Depletion: 2018. World Meteorological Organization, Global Ozone Research and Monitoring Project, Rep. 58, 588 pp.