HAL
open science

Updated validation of ACE and OSIRIS ozone and NO2 measurements in the Arctic using ground-based instruments at Eureka, Canada

K. Bognar, X. Zhao, K. Strong, C. D. Boone, A. E Bourassa, D. A. Degenstein, J. R. Drummond, A. Duff, Florence Goutail, D. Griffin, et al.

To cite this version:

K. Bognar, X. Zhao, K. Strong, C. D. Boone, A. E Bourassa, et al.. Updated validation of ACE and OSIRIS ozone and NO2 measurements in the Arctic using ground-based instruments at Eureka, Canada. Journal of Quantitative Spectroscopy and Radiative Transfer, 2019, 238 (November), pp.art. 106571. 10.1016/j.jqsrt.2019.07.014 . insu-02182891

HAL Id: insu-02182891

https://insu.hal.science/insu-02182891v1

Submitted on 11 Sep 2019

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

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

Journal of Quantitative Spectroscopy & Radiative Transfer xxx (xxxx) xxx

Contents lists available at ScienceDirect

Journal of Quantitative Spectroscopy & Radiative Transfer

journal homepage: www.elsevier.com/locate/jqsrt

Updated validation of ACE and OSIRIS ozone and NO_2 measurements in the Arctic using ground-based instruments at Eureka, Canada

K. Bognar[a,*], X. Zhao[b], K. Strong[a,*], C.D. Boone[c], A.E. Bourassa[d], D.A. Degenstein[d], J.R. Drummond[e], A. Duff[f], F. Goutail[g], D. Griffin[a,1], P.S. Jeffery[a], E. Lutsch[a], G.L. Manney[h,i], C.T. McElroy[j], C.A. McLinden[b], L.F. Millán[k], A. Pazmino[g], C.E. Sioris[b], K.A. Walker[a], J. Zou[a]

[a] Department of Physics, University of Toronto, Toronto, ON, Canada
[b] Environment and Climate Change Canada, Toronto, ON, Canada
[c] Department of Chemistry, University of Waterloo, Waterloo, ON, Canada
[d] Institute of Space and Atmospheric Studies, Department of Physics, University of Saskatchewan, Saskatoon, SK, Canada
[e] Department of Physics and Atmospheric Science, Dalhousie University, Halifax, NS, Canada
[f] Department of Physics, Queens's University, Kingston, ON, Canada
[g] LATMOS/IPSL, UVSQ Université Paris-Saclay, Sorbonne Université, CNRS, Guyancourt, France
[h] NorthWest Research Associates, Socorro, NM, USA
[i] Department of Physics, New Mexico Institute of Mining and Technology, Socorro, NM, USA
[j] Department of Earth and Space Science and Engineering, York University, Toronto, ON, Canada
[k] Jet Propulsion Laboratory, California Institute of Technology, Pasadena, CA, USA

ARTICLE INFO

Article history:
Received 21 December 2018
Revised 10 June 2019
Accepted 11 July 2019
Available online xxx

Keywords:
Validation
ACE
OSIRIS
Ozone
NO_2
Arctic

ABSTRACT

This paper presents long-term intercomparisons (2003–2017) between ozone and NO_2 measured by the Optical Spectrograph and Infra-Red Imager System (OSIRIS) and the Atmospheric Chemistry Experiment (ACE) satellite instruments, and by ground-based instruments at the Polar Environment Atmospheric Research Laboratory (PEARL), near Eureka, Nunavut, Canada (80°N, 86°W). The ground-based instruments include four zenith-sky differential optical absorption spectroscopy (DOAS) instruments, two Fourier transform infrared (FTIR) spectrometers, and a Brewer spectrophotometer. Comparisons of 14–52 km ozone partial columns show good agreement between OSIRIS v5.10 and ACE-FTS v3.5/3.6 data (1.2%), while ACE-MAESTRO v3.13 ozone is smaller than the other two datasets by 6.7% and 5.9%, respectively. Satellite profiles were extended to the surface using ozonesonde data, and the resulting columns agree with the ground-based datasets with mean relative differences of 0.1–12.0%. For NO_2, 12–40 km partial columns from ACE-FTS v3.5/3.6 and 12–32 km partial columns from OSIRIS v6.0 (scaled to 40 km) agree with ground-based partial columns with mean relative differences of 0.7–33.2%. Dynamical coincidence criteria improved the ACE to ground-based FTIR ozone comparisons, while little to no improvements were seen for other instruments, and for NO_2. A $\pm 1^\circ$ latitude criterion modestly improved the spring and fall NO_2 comparisons. The results of this study are consistent with previous validation exercises. In addition, there are no significant drifts between the satellite datasets, or between the satellites and the ground-based measurements, indicating that the OSIRIS and ACE instruments continue to perform well.

© 2019 Published by Elsevier Ltd.

1. Introduction

Long-term satellite datasets are essential to monitoring changes in the stratosphere. To ensure that the satellite measurements are well characterized, ground-based validation is required throughout the lifetime of the satellite instruments. This task is particularly challenging for satellites in high-inclination orbits, since they collect a large portion of their data in the Arctic, where the coverage of ground-based instruments is sparse. The Optical Spectrograph and InfraRed Imager System (OSIRIS) and the Atmospheric Chemistry Experiment (ACE) satellite instruments have been taking measurements in high-inclination orbits since 2001 and 2003, respectively. The ozone and NO_2 products from these instruments have

* Corresponding authors.
E-mail addresses: kbognar@atmosp.physics.utoronto.ca (K. Bognar), strong@atmosp.physics.utoronto.ca (K. Strong).
[1] Now at Environment and Climate Change Canada, Toronto, ON, Canada

https://doi.org/10.1016/j.jqsrt.2019.07.014
0022-4073/© 2019 Published by Elsevier Ltd.

Please cite this article as: K. Bognar, X. Zhao and K. Strong et al., Updated validation of ACE and OSIRIS ozone and NO_2 measurements in the Arctic using ground-based instruments at Eureka, Canada, Journal of Quantitative Spectroscopy & Radiative Transfer, https://doi.org/10.1016/j.jqsrt.2019.07.014

been validated before [1–5]. However, there are no recent comparisons in the Arctic involving OSIRIS and both ACE instruments. As the satellite data processing improves and new versions of the datasets are released, it is important to verify the consistency of ozone and NO_2 measurements at high latitudes. This task is especially important given that OSIRIS and ACE are currently the only satellite instruments measuring NO_2 profiles in the high Arctic.

Comparison of satellite and ground-based datasets in the high Arctic is challenging. Passive measurements are restricted to the sunlit part of the year, while the large solar zenith angles (SZAs) and small SZA variations pose challenges for both direct-sun and scattered-light instruments. Polar sunrise and sunset create conditions that lead to highly inhomogeneous stratospheric NO_2, while springtime comparisons are affected by the location of the polar vortex. When the polar vortex is strong, it isolates the airmass inside the core and hinders mixing with mid-latitude air. Substantially different trace gas concentrations inside and outside the polar vortex lead to strong gradients across the vortex boundary. Measurements taken in the vicinity of the polar vortex therefore need to be compared with care to account for the spatial variability of ozone and NO_2.

In addition to the atmospheric conditions, the harsh Arctic environment and logistical challenges restrict ground-based measurements to a few well-equipped stations. The Polar Environment Atmospheric Research Laboratory (PEARL) [6], located in Eureka, Canada (80°N, 86°W) is well suited to validate satellite instruments. PEARL is a collection of three separate facilities operated by the Canadian Network for the Detection of Atmospheric Change (CANDAC) since 2005. All but one of the ground-based instruments included in this study are located in the PEARL Ridge Lab (known as the Arctic Stratospheric Ozone Observatory prior to 2005), a facility 610 m above sea level and 15 km from the Environment and Climate Change Canada (ECCC) Eureka Weather Station (EWS).

PEARL and EWS host a large array of remote-sensing instrumentation, including radars, lidars, radiometers, and spectrometers covering the UV, visible, infrared, and microwave. At the PEARL Ridge Lab, ozone and NO_2 measurements have been made by zenith-scattered-light differential optical absorption spectroscopy (ZSL-DOAS) instruments on a campaign basis since 1999 (and year-round for 2007–2017), and by Fourier transform infrared (FTIR) spectrometers for 2006–2017 (year-round). In addition, ECCC Brewer spectrophotometers have been measuring ozone from 2004 to 2017. To support validation efforts, and to facilitate additional springtime measurements, Eureka has been the site for the annual Canadian Arctic ACE/OSIRIS Validation Campaigns since 2004 [7]. Ozone and NO_2 measurements have been used to validate the ACE and OSIRIS satellite instruments in a series of papers [7–14]. The PEARL facility is part of the Network for the Detection of Atmospheric Composition Change (NDACC), a network of more than 70 remote sensing stations around the globe that aim to monitor stratospheric and tropospheric changes and trends. The ZSL-DOAS and Bruker FTIR instruments follow standards and best practices outlined by the relevant working groups within NDACC, and data are submitted in a standardized format to the NDACC database.

This paper presents intercomparisons of ozone and NO_2 measurements from ground-based and satellite-borne instruments near Eureka, Canada, in the 2003–2017 period. Section 2 describes the instruments and datasets used in this study. The retrieval details for the ground-based ZSL-DOAS and FTIR instruments are given in Section 3. The comparison methodology and the details of the satellite partial columns, as well as the challenges presented by the diurnal variation of NO_2 are explained in Section 4. Comparison results between satellite instruments, and between satellite and ground-based instruments are presented in Section 5 for ozone and in Section 6 for NO_2. Section 7 examines the impact of the polar vortex in the spring, and the effect of clouds on ZSL-DOAS comparisons. Conclusions are given in Section 8.

2. Instruments

The ozone and NO_2 datasets used in this study, along with the corresponding abbreviations and temporal coverage, are listed in Table 1. Uncertainties, as reported in the datasets, are given in Table 2.

2.1. GBS ZSL-DOAS instruments

The University of Toronto Ground-Based Spectrometer (UT-GBS) and the PEARL-GBS [15] are both Triax-180 spectrometers from Jobin-Yvon/Horiba. The Triax-180 is a crossed Czerny-Turner imaging spectrometer with a grating turret that allows the selection of three resolutions and wavelength ranges. The UT-GBS and the PEARL-GBS differ in their input optics, gratings, and charge-coupled device (CCD) detectors. The UT-GBS took springtime measurements at the PEARL Ridge Lab from 1999 to 2001, 2003–2007, and 2009, while year-round measurements (with the exception of polar night) were taken in 2008 and 2010–2017. The UT-GBS was installed outside for 1999–2001, and it has been operating inside under a viewing hatch since 2003. In 2015, the instrument was placed in a temperature-controlled box to reduce the effect of temperature fluctuations in the lab. The PEARL-GBS was installed indoors in the PEARL Ridge Lab in 2006, and has been taking year-round measurements since then. The PEARL-GBS was set up in a temperature-controlled box in 2013, 2014, and 2017.

From 1999 to 2004, the UT-GBS used a thermoelectrically cooled CCD (230–250 K) with 2000×800 pixels (averaged across the 800 rows). The CCD was replaced in 2005 with a back-illuminated 2048×512 pixel CCD which operates at 201 K. The PEARL-GBS CCD is a newer version of the UT-GBS CCD and it includes a UV-enhanced coating on the CCD chip. The resolution in the trace gas retrieval windows varies across the measurement period based on the grating and slit selection, as well as the position of the CCD in each instrument. The typical resolution is 0.8–1.2 nm for ozone (up to 2.5 nm prior to 2005), 0.8–1.2 nm for NO_2 in the visible region (NO_2-vis), and 0.2–0.5 nm for NO_2 in the UV (NO_2-UV). The instruments have a field-of-view of approximately 1°. Since the two instruments are very similar and their ozone, NO_2-vis, and NO_2-UV data agree within 1%, the three pairs of datasets have been merged to create GBS time series. Twilight data were averaged when both instruments had measurements. Details of the data analysis can be found in Section 3.1. Both the UT-GBS and the PEARL-GBS are NDACC instruments, and data retrieved from the measurements are submitted to the NDACC database.

2.2. SAOZ ZSL-DOAS instruments

The Systéme d'Analyse par Observation Zénithale (SAOZ) instruments [16] form a global network that measures stratospheric trace gases using ZSL-DOAS. SAOZ instruments were deployed at the PEARL Ridge Lab in 2005–2017 as part of the Canadian Arctic ACE/OSIRIS Validation Campaigns. SAOZ-15 took springtime measurements in 2005–2009, while SAOZ-7 was installed in 2010 and took year-round measurements in 2011 and 2015–2017 with springtime data in the intervening years. For 2005–2007 and 2010, the instruments recorded spectra from inside the lab through a UV-transparent window. For 2008–2009 and since 2011, SAOZ was located in a box on the roof of the PEARL Ridge Lab.

Please cite this article as: K. Bognar, X. Zhao and K. Strong et al., Updated validation of ACE and OSIRIS ozone and NO_2 measurements in the Arctic using ground-based instruments at Eureka, Canada, Journal of Quantitative Spectroscopy & Radiative Transfer, https://doi.org/10.1016/j.jqsrt.2019.07.014

Table 1
Data products used in this study. The abbreviations listed are used in all subsequent figures and tables. The measurement periods are separated as spring only (S), spring and fall (S/F) and year-round (Y).

Data product	Abbreviation	Ozone	NO_2
GBS-vis	GV	S: 2003–2005 Y: Aug. 2006–2017	S: 2003–2005 Y: Aug. 2006–2017
GBS-UV	GU	–	S: 2007, 2009–2013, 2016 Y: 2008, 2014, 2015, 2017
SAOZ	SA	S: 2005–2010, 2012–2014 S/F: 2011, 2015–2017	S: 2005–2010, 2012–2014 S/F: 2011, 2015–2017
Bruker FTIR	BK	Y: Aug. 2006–2017	Y: Aug. 2006–2017
PARIS-IR	PA	S: 2006–2017	–
Brewer	BW	Y: 2004–2017	–
OSIRIS*	OS	Y: 2003–2017	Y: 2003–2017
ACE-FTS v3.5/3.6	AF	S/F: 2004–2017	S/F: 2004–2017
ACE-MAESTRO v3.13	AM	S/F: 2004–2017	–

* Data versions are v5.10 for ozone and v6.0 for NO_2.

Table 2
Reported uncertainty budgets for each of the datasets used in this study. Square brackets denote partial columns. For the list of abbreviations, see Table 1.

Instruments	Ozone		NO_2	
	DU	%	molec/cm^2	%
GV	22.7	6.6	[5.9×10^{14}]	[19.0]
GU	–		[6.5×10^{14}]	[22.8]
SA	23.4	5.9	[2.8×10^{14}]	[13.6]
BK	21.8	5.6	[2.3×10^{14}]	[7.5]
PA	21.9	4.9	–	
BW	1.3[a]	0.4[a]	–	
OS	[1.8][a]	[0.6][a]	[4.5×10^{13}][a,b]	[1.7][a,b]
AF	[1.1][a]	[0.4][a]	[1.8×10^{13}][a]	[1.1][a]
AM	[2.1][a,c]	[0.7][a,c]	–	

[a] Random uncertainties only.
[b] Based on estimate of uniform 1×10^8 molec/cm^3 uncertainty for each profile.
[c] Calculated using only the uncertainty values less than 10% to exclude profiles where the error calculation failed.

The SAOZ instruments are UV-visible spectrometers with a fixed grating that allows measurements in the 270–620 nm region. Spectra are recorded with an uncooled 1024-pixel linear photodiode array detector. The resolution is approximately 1 nm across the detector, and the instruments have a field-of-view of 20°. SAOZ-15 and SAOZ-7 are identical instruments and show excellent agreement, therefore measurements from the two instruments are treated as a single dataset. Details of the data analysis are described in Section 3.1. While SAOZ instruments are NDACC certified, the Eureka instruments are not part of the NDACC network. The SAOZ V3 dataset was used in this study. Changes compared to the V2 dataset are described in Section 3.1.

2.3. CANDAC Bruker FTIR

The CANDAC Bruker IFS 125HR Fourier transform infrared spectrometer was installed in the PEARL Ridge Lab in 2006 [17]. Solar absorption spectra are recorded using either a mercury cadmium telluride (HgCdTe) or an indium antimonide (InSb) detector (both liquid-nitrogen-cooled), and a potassium bromide (KBr) beamsplitter. Seven narrow-band interference filters are used to cover a range of 600–4300 cm^{-1}. Measurements take approximately 4–8 min, consist of two to four co-added spectra, and have a resolution of 0.0035 cm^{-1}. No apodization is applied to the measurements. The Bruker FTIR is part of NDACC, and retrieved ozone profiles are submitted to the NDACC database, while the NO_2 retrievals are currently a research product. The retrieval details for both ozone and NO_2 can be found in Section 3.2.

2.4. PARIS-IR

The Portable Atmospheric Research Interferometric Spectrometer for the InfraRed (PARIS-IR) took measurements at the PEARL Ridge Lab in 2004–2017 as part of the Canadian Arctic ACE/OSIRIS Validation Campaigns. Measurements are only included for the 2006–2017 period, as the instrument has been operated in a consistent fashion since the 2006 campaign. PARIS-IR has a design similar to that of the ACE Fourier Transform Spectrometer (ACE-FTS) [18]. Solar absorption spectra are recorded using liquid-nitrogen-cooled HgCdTe and InSb detectors, and a zinc selenide (ZnSe) beamsplitter. The measurements are recorded in the 750–4400 cm^{-1} range, at a 0.02 cm^{-1} resolution and without the use of narrow-band filters. Measurements are recorded approximately every 7 min and consist of 20 co-added spectra. No apodization is applied to the measurements. The details of the ozone retrieval can be found in Section 3.2.

2.5. Brewer spectrophotometer

Brewer instruments use a grating with a slit mask to measure the intensity of direct sunlight at six wavelengths in the UV range [19]. The first two wavelengths are used for internal calibration and SO_2 retrievals, respectively. Ozone total columns are calculated using relative intensities at the four remaining wavelengths (310.1, 313.5, 316.8, and 320 nm), with slight changes to the analysis to account for the high latitude of the measurement site [8]. Several Brewer instruments were deployed in Eureka over the 2003–2017 period. In this study, only Brewer #69 (a MKV single monochromator) is included, since this instrument measured hourly ozone for 2004–2017. During this time, Brewer #69 was located on the roof of the EWS building.

2.6. Ozonesondes

Electrochemical concentration cell (ECC) ozonesondes are launched by ECCC from the Eureka Weather Station on a weekly basis [20]. During the intensive phase of the Canadian Arctic ACE/OSIRIS Validation Campaigns (2004–2017, typically early March), ozonesondes were launched daily, weather permitting. In this study, ozonesondes were used in the ZSL-DOAS retrievals (Section 3.1), to extend satellite partial columns of ozone to the surface (Section 4.3), and to initialize the photochemical box model used for NO_2 diurnal scaling (Section 4.3).

2.7. OSIRIS

The Odin satellite, carrying the OSIRIS instrument [21,22], was launched in February 2001. OSIRIS measures limb-radiance profiles at a 1–2 km resolution, and measurements near PEARL are available throughout the sunlit part of the year. The optical spectrograph in OSIRIS is a UV-visible grating spectrometer that measures scattered sunlight from 280 to 800 nm with 1 nm resolution. Spectra are recorded on a 1353 × 286 pixel CCD detector.

The ozone profiles in the version 5.10 dataset [23] used in this study are retrieved using the SaskMART algorithm. SaskMART [2] is a multiplicative algebraic reconstruction technique (MART) that uses information from the UV and visible ozone absorption bands. The SASKTRAN radiative transfer model [24] is used as the forward model in the retrievals. The v5.10 dataset corrects a pointing bias drift, apparent in the preceding version from 2012 onward. The retrieval algorithm is unchanged compared to previous versions. The NO_2 retrievals use a different approach. A modified DOAS algorithm is used to retrieve slant column densities (SCD), and the SCDs are converted to profiles using MART and the SASKTRAN model. The OSIRIS version 6.0 NO_2 [5] is used in this study. The v6.0 dataset is substantially different from the previous operational product (v3.0) which used optimal estimation and a different forward model.

2.8. ACE-FTS and ACE-MAESTRO

ACE [25], on board the SCISAT satellite, consists of two main instruments: the Fourier Transform Spectrometer (ACE-FTS) and the Measurement of Aerosol Extinction in the Stratosphere and Troposphere Retrieved by Occultation (ACE-MAESTRO). Launched in August 2003, SCISAT takes solar occultation measurements. The instruments collect data near PEARL during sunset from late February to mid-March, and during sunrise from late September to mid-October.

The ACE-FTS is a high-resolution (0.02 cm^{-1}) infrared Fourier transform spectrometer that measures in the 750–4400 cm^{-1} range. Interferograms are recorded on two photovoltaic detectors (InSb and HgCdTe). The first step in the retrieval is the determination of pressure and temperature profiles based on a detailed CO_2 analysis. The volume mixing ratio (VMR) profiles are then retrieved using a global nonlinear least squares fitting algorithm [26]. The ACE-FTS data version 3.5/3.6 [27] is included in this study. The v3.5 and v3.6 data use identical algorithms in different computing environments. The current processing differs from the previous version (v3.0) only in the low-altitude pressure and temperature inputs from October 2011 onward.

The ACE-MAESTRO is a UV-visible-near-IR double spectrograph with a resolution of 1–2 nm. The two channels cover 280–550 nm and 500–1030 nm, and spectra are recorded on 2014-pixel linear photodiode array detectors. Profiles are retrieved using a two-step approach where SCDs are retrieved using a modified DOAS procedure, and vertical profiles are derived using a nonlinear Chahine relaxation inversion [28]. The retrievals use ACE-FTS temperature and pressure profiles. The ACE-MAESTRO version 3.13 ozone product is used in this study. The v3.13 retrieval improves the reference spectrum and error calculations of the preceding version (v3.12/3.12.1). The v3.13 dataset does not include NO_2, since it is retrieved from the UV spectrometer, and the UV channel has been experiencing gradual degradation since the launch. UV data are not considered useful past October 2010[2] ACE-MAESTRO NO_2 was excluded from this study due to the low coincidence count of the available data in the v3.12.1 dataset.

[2] ACE-MAESTRO Level 2 Version 3.13 Data Description and File Formats, https://databace.scisat.ca/level2/mae_v3.13/ACE-MAESTRO-V3.13-Data.pdf. Accessed 2018/09/28.

3. Data analysis for ground-based instruments

3.1. ZSL-DOAS measurements

The GBS and SAOZ instruments use the DOAS technique [29] to retrieve ozone and NO_2 columns from zenith-scattered sunlight. The GBS and SAOZ analyses were performed independently, with slight differences in the retrieval settings.

The main product of DOAS is the differential slant column density (dSCD), the amount of trace gas in the slant column minus the amount in a reference spectrum. The GBS dSCDs were retrieved with daily reference spectra, while the SAOZ retrievals used a fixed reference spectrum for each year. The dSCDs were retrieved using the settings recommended by the NDACC UV-visible Working Group [30]. For ozone, SAOZ retrievals used the recommended 450–550 nm window, while the GBS instruments used 450–545 nm to avoid irregularities at the CCD edge. For NO_2, the GBS-vis datasets used the recommended 425–490 nm window, while the SAOZ retrievals used an extended, range, 410–530 nm. The GBS-UV dataset used the 350–380 nm window. The NO_2-UV data are not a standard NDACC product, but the retrievals followed the NO_2-vis recommendations as closely as possible.

For each twilight, dSCDs in the 86–91° SZA range were used in the vertical column density (VCD) retrieval. Reference column densities (RCDs) were calculated using the Langley plot method. For the GBS instruments, daily RCDs were calculated from the average of the RCDs for each twilight. For SAOZ, a fixed RCD was calculated for each year, since yearly references were used in the DOAS analysis. Single VCD values for each twilight were calculated as the mean of the individual vertical columns in the given SZA range, weighted by the DOAS fitting error, divided by the air mass factor (AMF).

The AMFs used in the VCD retrieval were provided by NDACC in the form of look-up tables [30]. The ozone AMF calculations require the input of daily ozone data. The GBS retrievals used total columns interpolated from ozonesonde data, while the SAOZ analysis used measured slant column densities. The NO_2 AMF look-up tables, compiled separately for sunrise and sunset conditions, do not require prior vertical column information. The NO_2 concentration below 12 km and above 60 km in the look-up tables is set to zero, and so the ZSL-DOAS NO_2 VCDs in this study are 12–60 km partial columns.

ZSL-DOAS measurements are particularly challenging in the high Arctic. The ideal SZA window of 86–91° is not available for much of the sunlit part of the year, and the maximum SZA at the summer solstice is just over 76°. The SAOZ VCDs are only retrieved in the spring and fall, when the 86–91° window is available. In order to extend the measurements into the polar day, the GBS retrievals use the highest available 5° SZA window in the summer. Around the summer solstice, however, the maximum AMFs for both ozone and NO_2 are only about one fourth of the AMFs at 90° SZA. In addition, the range in AMFs for SZAs of 71–76° is smaller than 1, while the AMF range is greater than 10 for the NDACC recommended SZA window. This leads to larger uncertainties in the summertime VCD retrievals. Spring and fall present their own unique challenges. The lack of high-sun spectra to use as daily references negatively impacts the quality of the GBS dSCDs, and small NO_2 concentrations lead to very large uncertainties in the GBS RCD calculations.

The GBS uncertainty calculations follow Table 4 of Hendrick et al. [30], with updated values to more accurately reflect the GBS retrievals. The mean total uncertainty for the 2003–2017 GBS ozone dataset was calculated to be 6.6%, which is larger than the

5.9% reported for NDACC ozone columns [30]. The larger value, however, is consistent with the challenges of high-latitude measurements outlined above. The GBS NO_2-vis and NO_2-UV datasets have mean total uncertainties of 19.0% and 22.8 %, respectively. To ensure the consistency of the daily RCD and uncertainty calculations, GBS VCDs were only computed if both twilights had measurements. The SAOZ dataset contains only the errors from the DOAS fitting procedure. The total uncertainty of SAOZ ozone was estimated to be 5.9% [30]. SAOZ NO_2 measurements have an estimated precision of 1.5×10^{14} molec/cm^2 and accuracy of 10%. Combined in quadrature, this yields a 13.6% total uncertainty for the SAOZ NO_2 measurements used in this study.

The SAOZ V3 dataset is different from the V2 data used in previous validation studies. For ozone, the changes are limited to new reference spectra (and therefore reprocessed dSCDs and new RCD values) for 2008–2010. For NO_2, the changes are more substantial. The V2 dataset was processed using a single set of AMFs representative of Arctic summer evenings, and the retrievals produced total columns. The V3 retrievals use the NDACC AMF look-up tables, and produce 12–60 km partial columns. The same wavelength range (410–530 nm) was used for both NO_2 retrievals.

To investigate the differences between satellite minus GBS and satellite minus SAOZ intercomparisons, we retrieved ozone and NO_2 VCDs from the original SAOZ dSCDs using the GBS VCD retrieval code. This retrieval extended the SAOZ data to include year-round measurements in 2011 and 2015–2017. This dataset (hereafter $SAOZ_{allyear}$) used the same settings as the SAOZ retrieval, with the exception of the SZA range. Similar to the GBS retrievals, the highest available 5° SZA window was used to obtain summer data.

3.2. FTIR measurements

The Bruker FTIR and the PARIS-IR employ a similar technique to retrieve vertical VMR profiles from measured solar-absorption spectra. VMR profiles are retrieved using the SFIT4 version 0.9.4.4 retrieval algorithm, which, as with the previous SFIT2 retrieval algorithm, is based upon the methods of Pougatchev et al. [31]. SFIT4 uses an optimal estimation method that iteratively adjusts the retrieved VMR to best fit the measured spectra [32]. The trace gas a priori profiles required by SFIT4 are provided by the mean of a 40-year (1980–2020) run of the Whole Atmosphere Community Climate Model (WACCMv4) [33], while daily pressure and temperature profiles used in the retrieval are provided by the U.S. National Centers for Environmental Prediction (NCEP) and interpolated to the geolocation of PEARL. Spectroscopic line lists are from HITRAN 2008 [34] as recommended by the NDACC Infrared Working Group (IRWG).

The ozone retrievals for both instruments use a single microwindow, spanning 1000.0–1004.5 cm^{-1} [12], which also contains the interfering species H_2O, CO_2, and the ozone isotopologues O_3^{668} and O_3^{686}. Profiles are simultaneously retrieved for H_2O and CO_2 from the Bruker FTIR spectra, whereas for PARIS-IR spectra H_2O and the ozone isotopologues are retrieved as profiles. Profiles of the remaining species, O_3^{668} and O_3^{686} for the Bruker FTIR and CO_2 for PARIS-IR, are scaled from their a priori values. Retrievals are performed on a 29-layer grid, from 0.61 to 100 km, for PARIS-IR, and on a 47-layer grid, from 0.61 to 120 km, for the Bruker FTIR.

The a priori covariance matrix for the Bruker FTIR ozone retrievals is formed from diagonal values of 5% from the surface (0.61 km) to approximately 45 km. Above 45 km, the diagonal values are scaled to 4.2% to reduce oscillations in the retrieved profiles. Off-diagonal elements are formed from an exponential inter-layer correlation, with a correlation width of 2 km, applied from the surface to the top of the atmosphere at 120 km. The a priori covariance matrices of the interfering species H_2O and CO_2 are formed with diagonal elements of 20% for all altitudes with no inter-layer correlation. These a priori covariance matrices provided the optimal degrees of freedom for signal (DOFS) while minimizing unphysical oscillations in the retrievals. The mean DOFS for ozone is approximately 5, with minimum values near 4 and maximum values near 6.

The a priori covariance matrix for PARIS-IR is constructed from diagonal values of 7% for all altitudes, with no inter-layer correlation. The a priori covariance matrices of the interfering species H_2O, O_3^{668} and O_3^{686} are formed with diagonal elements of 20% for all altitudes again with no inter-layer correlation. The mean DOFS for ozone is approximately 3, with minimum values of approximately 1 and maximum values around 4.5.

The Bruker FTIR NO_2 retrievals use five microwindows centered on 2914.65, 2918.23, 2919.53, 2922.58, and 2924.84 cm^{-1}. The interfering species are CH_4, CH_3D, H_2O, ozone and OCS. CH_4 and CH_3D are retrieved as profiles, whereas H_2O, ozone, and OCS are scaled from their a priori values. The retrievals are performed on the same 47-level grid as for ozone. The a priori covariance matrix for the NO_2 retrieval is formed from diagonal values of 40% for all altitudes, and an exponential inter-layer correlation (with a correlation width of 4 km) for the off-diagonal elements. The a priori covariance matrices of the interfering species CH_4 and CH_3D are formed with diagonal elements of 25% for all altitude levels, with no inter-layer correlation. The mean DOFS for the NO_2 retrieval is 1.2, with minimum values near 0.8 and maximum values around 1.6. The DOFS show strong seasonality, with spring and fall values between 1.2 and 1.6, and summertime values of 1-1.2.

A full error analysis was performed following Rodgers [32], which includes the forward model parameter error and the measurement noise error. Adding these in quadrature, the mean uncertainty for the entire ozone time series from 2006 to 2017 is 5.6% of the retrieved total column for the Bruker 125HR and 4.9% for PARIS-IR. These values are similar to mean uncertainties of other FTIR ozone retrievals from the NDACC IRWG. The mean uncertainty for 2006–2017 is 7.5% for the Bruker FTIR NO_2 retrievals. The smoothing error was not included in the mean uncertainty calculations [35].

The retrievals were quality controlled using the root-mean-squared (RMS) values of the residual and the DOFS. An RMS:DOFS ratio of 1.0 was used in the Bruker FTIR ozone retrieval, while the PARIS-IR retrieval used a value of 6.0, and the Bruker FTIR NO_2 retrieval used a value of 1.5. Profiles with RMS:DOFS ratios higher than the aforementioned limits were excluded to eliminate poor spectral fits and maintain adequate retrieved information. Additionally, several outliers were omitted from the datasets based on a qualitative analysis of the fitted spectra.

4. Comparison methodology

The validation metrics used to assess the similarity of the datasets are described in Section 4.1. Coincident measurements used for the comparisons were selected using the methods outlined in Section 4.2. The procedures for extending ozone profiles using ozonesonde data, and for scaling NO_2 columns using a photochemical model are described in Section 4.3. The methodology used to assess the long-term consistency of the satellite datasets is described in Section 4.4.

4.1. Comparison metrics

To evaluate systematic differences between the datasets, mean absolute and relative differences were used. The mean absolute difference between a set of coincident measurements x and y is given

by

$$\Delta_{abs} = \frac{1}{N}\sum_{i=1}^{N}(x_i - y_i), \tag{1}$$

where N is the number of coincident measurements. The mean relative difference, defined with respect to the average of the measurement pairs, is given by

$$\Delta_{rel} = \frac{1}{N}\sum_{i=1}^{N}\frac{(x_i - y_i)}{(x_i + y_i)/2} \times 100\%. \tag{2}$$

The standard errors ($\sigma/\sqrt{N}$, where σ is the standard deviation of the differences) were also calculated for the mean absolute and relative differences. The standard error is the reported error throughout this paper. In addition, to quantify the statistical spread of the absolute differences, the root-mean-square deviation (*RMSD*) is used:

$$RMSD = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(x_i - y_i)^2}. \tag{3}$$

Unlike the standard deviation of the differences, *RMSD* captures the bias between the datasets as well. If there is no bias between the datasets, then $RMSD = \sigma$. For satellite to ground-based comparisons, we use the sign convention such that x is the satellite dataset and y is the ground-based dataset.

The statistical dependency of the datasets was evaluated using Pearson's correlation coefficient (R). In correlation plots, the linear relationship between the datasets was characterized using the ordinary least squares (OLS) method, and the reduced major-axis (RMA) method [36]. The RMA solution is equivalent to minimizing the sum of squares of the perpendicular distances between the points and the fitted line. Since the RMA solution is symmetrical, it doesn't require the assignment of one dataset as the independent variable. Measurement uncertainties were not included in the linear fits, since some of the datasets include random errors only, while some of the datasets do not provide uncertainty values for individual measurements, only an estimate of the overall uncertainty.

Since pairwise comparison metrics are sensitive to uncertainties in both datasets, we use triple colocation analysis (TCA), a method commonly used for global validation studies [37–43], to estimate uncertainties in the individual datasets. By adding a third coincident dataset, TCA allows an estimate of the root-mean-square-error (*RMSE*) and correlation with respect to the unknown truth for each dataset. The *RMSE* is the square root of the random error variance, and is given by

$$RMSE(x) = \sqrt{\sigma_x^2 - \frac{\sigma_{xy}\sigma_{xz}}{\sigma_{yz}}}, \tag{4}$$

for one dataset, using the three coincident datasets x, y and z. σ_{xy}, σ_{xz}, σ_{yz} are the covariances of the datasets, and σ_x^2 is the variance of the measurements in question. The correlation with respect to the unknown truth is defined as

$$R^t = \sqrt{\frac{\sigma_{xy}\sigma_{xz}}{\sigma_x^2\sigma_{yz}}}. \tag{5}$$

The *RMSE* and R^t are analogous to the *RMSD* and R values from pairwise comparisons, however while *RMSD* and R are sensitive to uncertainties in both datasets, *RMSE* and R^t are only sensitive to uncertainties in dataset x.

All comparison metrics (pairwise or triple colocation) used in this study are affected by colocation mismatch, that is differences between the spatiotemporal sampling of the inhomogeneous ozone and NO_2 distributions by different instruments. Ozone colocation errors have been estimated by Verhoelst et al. [44]. They used GOME-2 and NDACC ozone measurements, combined with modeled observations, to quantify the error budgets of satellite to ground-based intercomparisons for a host of ground-based stations (67°N to 75°S). They found that colocation errors dominate the error budgets, and can account for differences of 10% or more at high-latitude stations. Using similar methods, colocation errors between OSIRIS and ACE-FTS ozone can also be estimated. For the coincidence criteria used in this study (12 h and 500 km), and including the Arctic (poleward of 60°N) only, the mean relative difference between OSIRIS and ACE-FTS 10–55 km partial columns is expected to be 6.4–6.9%[3]. Colocation errors for satellite to ground-based comparisons are expected to be similar, while for NO_2, the values are expected to be larger due to the high latitudinal gradient and diurnal variation.

The contribution of colocation error to the *RMSE* values varies depending on the combination of instruments, due to differences in viewing geometries and measurement techniques. In order to limit the effect of colocation error, the calculated *RMSE* values are specific to instrument pairs, and only the sum of the *RMSE* values is reported for each pair. This way, satellite datasets are not penalized when grouped with two ground-based instruments, and vice versa. *RMSE* values for the individual instruments were calculated as the average *RMSE* from all triplets that included both instruments in the pair. For example, using ACE-FTS and GBS ozone, the triplets with SAOZ, Bruker FTIR, PARIS-IR, ACE-MAESTRO, and OSIRIS data were considered, the *RMSE* values (five for both ACE-FTS and GBS) were averaged, and then added to get the final value shown in Table 3. This process was repeated for all instrument pairs considered in this study. The final *RMSE* values provide an upper limit on the expected spread between data from various instrument pairs. R^t values for each instrument were calculated in a similar fashion, except those values were not added in the final step.

Throughout this paper, the convention is that 'spring' and 'fall' are defined as the periods when the sun crosses the horizon daily (i.e. 90° SZA is available). These periods, from day 53 to day 105 (February 23 to April 14/15) and from day 240 to day 291 (August 27/28 to October 17/18), include all ACE measurements, and all ZSL-DOAS measurements with the ideal 86–91° SZA range. The remainder of the sunlit part of the year is referred to as summer.

4.2. Coincidence criteria

Temporal coincidence criteria were selected based on the measurement methods of the instruments. For twilight-measuring instruments (ACE-FTS, ACE-MAESTRO, and the ZSL-DOAS instruments), comparisons were restricted to the same twilight. In addition, comparisons between ACE-FTS and ACE-MAESTRO were restricted to the same occultation. For all other instrument pairs, coincidences were generated by pairing measurements from both datasets to the nearest measurement in the other dataset, within a ±12 h time window. For triple colocation, these coincidence criteria were applied simultaneously to all three pairs within the group.

For spatial coincidences, satellite measurements within 500 km of the PEARL Ridge Lab were considered. The approximate location of the air masses sampled by each instrument is shown in Fig. 2 of Adams et al. [8]. The primary reason for using a 500 km radius was to reduce the impact of the spring and fall latitudinal NO_2 gradient on the comparison results. These impacts are assessed in Section 7.1. Comparison results for a 1000 km radius around PEARL show that for NO_2, mean differences change significantly and the correlation coefficients decrease, when compared to the 500 km results. Fig. 5 of Adams et al. [8] shows modeled ratios of NO_2 par-

[3] P. Sheese, personal communication.

Table 3
Sum of the averaged *RMSE* values for all possible instrument pairs that involve at least one satellite instrument. The values were calculated using TCA, as described in Section 4.1. The uncertainty values are the standard errors on the averaged *RMSE* values, combined in quadrature. The number of triplets considered in the average (i.e. the number of third instruments), as well as the total number of triple coincidences (N_{tot}) are indicated for each pair. Instrument abbreviations are given in Table 1.

Ozone				NO_2			
Instrument Pair	Sum of *RMSE* (DU)	Triplets	N_{tot}	Instrument Pair	Sum of *RMSE* ($x10^{14}$ molec/cm^2)	Triplets	N_{tot}
OS, AF	25.8 ± 2.8	5	4047		–		
OS, AM	35.2 ± 1.4	5	3550		–		
AF, AM	21.2 ± 1.7	5	5229		–		
OS, GV	36.1 ± 2.1	6	23,303	OS, GV	5.7 ± 0.5	3	1918
OS, SA	32.2 ± 3.0	6	9687	OS, GU	5.6 ± 0.3	3	1749
OS, BK	36.0 ± 4.5	6	23,309	OS, SA	5.5 ± 0.4	3	1204
OS, PA	42.2 ± 3.4	6	17,268	OS, BK	4.7 ± 0.3	3	2597
OS, BW	27.2 ± 2.1	4	33,372	AF, GV	3.3 ± 0.4	3	887
AF, GV	38.4 ± 4.5	5	2874	AF, GU	3.5 ± 0.4	3	656
AF, SA	33.3 ± 5.5	5	3169	AF, SA	3.7 ± 0.3	3	925
AF, BK	41.1 ± 5.2	5	5943	AF, BK	3.7 ± 0.1	3	482
AF, PA	41.5 ± 5.7	5	12,252		–		
AM, GV	46.5 ± 3.3	5	2439		–		
AM, SA	40.0 ± 4.7	5	2669		–		
AM, BK	46.9 ± 5.2	5	5277		–		
AM, PA	50.7 ± 4.4	5	10,950		–		

Table 4
Drift values and corresponding uncertainties for the satellite minus ground-based daily mean relative difference time series, as described in Section 4.4. The variance-weighted mean value is also indicated for each satellite data product. Drifts that are significant based on the uncertainty alone are highlighted in bold. Whether these drifts are meaningful, or the results of evolving comparison statistics, is discussed in Sections 5.4 (for ozone) and 6.3 (for NO_2). None of the drifts are significant based on the number of years (n*) required to detect a real drift in the datasets. Instrument abbreviations are given in Table 1.

Satellite	Ground-based	Ozone drift (%/decade)		NO_2 drift (%/decade)	
Instrument	Instrument	Pairwise	Mean	Pairwise	Mean
OS	GV	−0.9 ± 3.1	**1.2 ± 0.9**	−2.9 ± 9.5	−5.1 ± 5.7
	GU	–		−1.2 ± 13.7	
	SA	−1.5 ± 2.7		−4.0 ± 13.8	
	BK	0.4 ± 2.2		**−10.4 ± 10.3**	
	PA	−2.3 ± 5.1		–	
	BW	**2.7 ± 1.3**		–	
AF	GV	−5.0 ± 5.1	**−3.3 ± 2.4**	7.4 ± 12.8	**8.3 ± 7.7**
	GU	–		5.3 ± 18.5	
	SA	−2.5 ± 4.2		12.8 ± 13.3	
	BK	−4.6 ± 4.9		3.6 ± 21.4	
	PA	−1.1 ± 5.5		–	
AM	GV	−2.3 ± 7.7	−0.9 ± 3.3	–	–
	SA	−0.2 ± 7.3		–	
	BK	−4.1 ± 6.8		–	
	PA	1.2 ± 5.3		–	

tial columns at various latitudes for SZA=90°, as a function of day of the year. Ratios of partial columns at 78°N over 82°N (typical difference for coincidences within 500 km) could be as high as 7 in early spring and late fall, while latitude differences typical for a 1000 km radius correspond to ratios of 20–25 during the same periods. Ozone comparisons show only small differences when the radius is increased to 1000 km. Using the 500 km radius ensures that the results are directly comparable to Adams et al. [8], who also used this radius around PEARL to compare datasets from the instruments included in this study.

4.3. Partial columns

ACE-FTS and ACE-MAESTRO VMR profiles were converted to number density using ACE-FTS pressure and temperature profiles. The OSIRIS profiles are reported as number densities. For the integration to partial columns, profiles were accepted only if all levels in the selected altitude range had valid values. While negative VMR values for ACE-FTS and ACE-MAESTRO were accepted as valid data, none of the profiles considered in the comparisons include negative values within (or immediately outside) the ozone or NO_2 partial column ranges.

For comparisons between satellite instruments, ozone partial columns from 14 to 52 km were calculated, in order to maximize the number of available profiles from all three satellite instruments. For comparison to ground-based instruments, the satellite partial columns were extended down to the altitude of the given instrument (610 m for the PEARL Ridge Lab and 10 m for the Eureka Weather Station; a difference of 1–2 DU) using ozonesonde profiles. This approach is similar to the methods of Adams et al.

[8] and Fraser et al. [9]. Sonde profiles were interpolated to the satellite measurement time, and the resulting profiles were smoothed between 12–16 km with a moving average to avoid discontinuities in the joint profile. Excluding the smoothing step results in a mean change of only 0.3% in the satellite total columns. Ozone above 52 km was neglected, since it accounts for less than 0.2% (< 1 DU) of the total ozone column according to the NDACC ozone climatology [30] for Eureka.

For NO_2 partial columns, an altitude range of 12–40 km was chosen. The lower altitude limit was determined by the ZSL-DOAS retrievals, since the standardized NDACC AMFs only include NO_2 above 12 km. The upper value was set to 40 km to ensure that the results are comparable to Adams et al. [8]. No correction was applied to extend the columns above 40 km, since NO_2 above that altitude accounts for less than 2% of the total column, which is much smaller than the measurement uncertainties for the ground-based instruments. For OSIRIS, the upper altitude limit was reduced to 32 km, since most profiles only extended to that altitude. For comparison to ground-based instruments, OSIRIS NO_2 partial columns were scaled to 40 km using NDACC look-up table profiles calculated using the time, geolocation, and mean wavelength of the OSIRIS measurements.

Diurnal variation of NO_2 must be considered when comparing measurements taken at different times of the day. In the spring and fall, NO_2 increases during the day due to release from nighttime reservoirs. During the polar day (mid-April to late-August), NO_2 decreases at noon due to photolysis to NO. To account for the diurnal variation, NO_2 partial columns were scaled to local noon [e.g. 8,13] using a photochemical box model [45,46]. The model was initialized for 80°N using the NDACC surface albedo climatology and ozonesonde profiles of ozone and temperature interpolated to local noon for each day. For a detailed discussion of the scaling procedure, see Adams et al. [8].

Diurnal variation of NO_2 also leads to errors in individual measurements through the so-called diurnal effect [47–50]. The diurnal effect occurs mainly because sunlight passes through a range of SZA before reaching the instruments, and NO_2 is at different stages of its diurnal cycle for different SZA. For ACE-FTS, NO_2 profiles below 25 km can increase by up to 50% as a result of the diurnal effect [4]. For OSIRIS, these errors are less relevant since only measurements with SZA greater than 85° are expected to change due to the diurnal effect [1,49], and the v6.0 dataset used here contains no such measurements near PEARL. The ZSL-DOAS instruments likely underestimate NO_2, since the SZA at 30 km along the estimated line-of-sight is ∼3° smaller (∼2° for UV) than the SZA at the instrument location for the standard 86–91° SZA window. Bruker FTIR measurements are affected in the early spring, when SZA in the 30 km layer can be up to 5° smaller than the SZA at the ground. The discrepancy for the Bruker FTIR, however, quickly decreases in the spring as the sun climbs higher in the sky. In addition to the diurnal effect, the diurnal variation of NO_2 also leads to strong latitudinal gradients in the spring and fall. NO_2 concentrations are smaller at higher latitudes, due to the decreasing number of daylight hours with increasing latitude. The impact of the diurnal effect and the latitudinal gradient on the comparison results is examined in Section 7.1.

4.4. Time series analysis

Given the long data record for all instruments included in this study (see Table 1), the decadal stability of the satellite data products can be assessed. For each instrument pair, the daily mean relative differences were calculated, and a linear fit with respect to time was used to obtain an estimate of the drift between the two instruments [e.g. 51,52]. The linear regression was performed using a bi-square weighted robust fitting method [53]. Robust methods are preferable over OLS methods, since the former are less sensitive to outliers and data gaps. The uncertainties given by the robust fit were verified using bootstrap resampling, [54] and the two uncertainty calculations were found to be in very good agreement.

The uncertainties reported for the drift values (σ) were calculated using a correction for the autocorrelation of the noise, [52,55] such that

$$\sigma = 2\sigma_{fit} \times \sqrt{\frac{1+\phi}{1-\phi}}, \tag{6}$$

where σ_{fit} is the uncertainty from the robust fit, and ϕ is the lag-1 autocorrelation of the noise. We take the residual daily mean relative differences to represent the distribution of noise in the data [56]. The values of σ yield a more conservative estimate of the uncertainty as compared to the fit uncertainties. Potential seasonality in the relative difference time series was not taken into account explicitly, due to the limitations of OLS fitting methods for sparsely sampled time series, and the large scatter (relative to the potential seasonality) in the relative difference datasets. To assess the feasibility of drift detection for each dataset, we calculated the number of years (n^*) required to detect a real drift of a given magnitude in the data, as given by Weatherhead et al. [55]:

$$n^* = \left(\frac{3.3\sigma_N}{|\,\omega\,|}\sqrt{\frac{1+\phi}{1-\phi}}\right)^{2/3}. \tag{7}$$

The factor of 3.3 returns n^* for the given drift value (ω) with 90% certainty, and σ_N is the standard deviation of the noise. The statistical significance of the drift value for each dataset was assessed using both the error on the drift (σ) and the number of years (n^*) required to detect the drift with 90% certainty.

In addition to the drift values for each satellite minus ground-based time series, the mean drift for each satellite data product was calculated using a variance-weighted mean [51]. Weights of σ_i^{-2} were used, where σ_i is the uncertainty of the drift value for the ith instrument pair in the average. The uncertainty on the mean drift is given by $(\sum \sigma_i^{-2})^{-1/2}$.

4.5. Averaging kernel smoothing

Satellite profiles were not smoothed in this study, for reasons detailed below. The OSIRIS, ACE-FTS and ACE-MAESTRO satellite instruments measure at a higher vertical resolution than the ground-based instruments considered here. To account for this difference, the satellite profiles might be smoothed with the ground-based averaging kernels according to the method of Rodgers and Connor [57]. Smoothing the satellite profiles for comparisons with the Bruker FTIR and the PARIS-IR is straightforward, and is routinely implemented in validation studies [e.g. [3,10,12]]. However, given the good sensitivity of the FTIR instruments to most of the ozone and NO_2 columns [13,17], smoothing is expected to have a small impact on ozone and NO_2 comparisons.

The Brewer and ZSL-DOAS retrievals, on the other hand, do not provide averaging kernels or use a priori profiles. To address this problem, approximate ZSL-DOAS averaging kernels were developed at the Belgian Institute for Space Aeronomy (BIRA-IASB) in the form of look-up tables. The averaging kernel calculations are based on the methods of Eskes and Boersma [58]. In the current iteration, however, the averaging kernels are calculated for 90° SZA only. This limits their use to spring and fall for PEARL data. Furthermore, most of the changes in the smoothed profiles can be attributed to the systematic differences between the unsmoothed satellite profiles and the climatology used as a priori in the smoothing process.

Considering only the profiles coincident with ground-based measurements, satellite-plus-sonde ozone columns change, on average, by less than 0.2% and 1.4% when smoothed with the Bruker

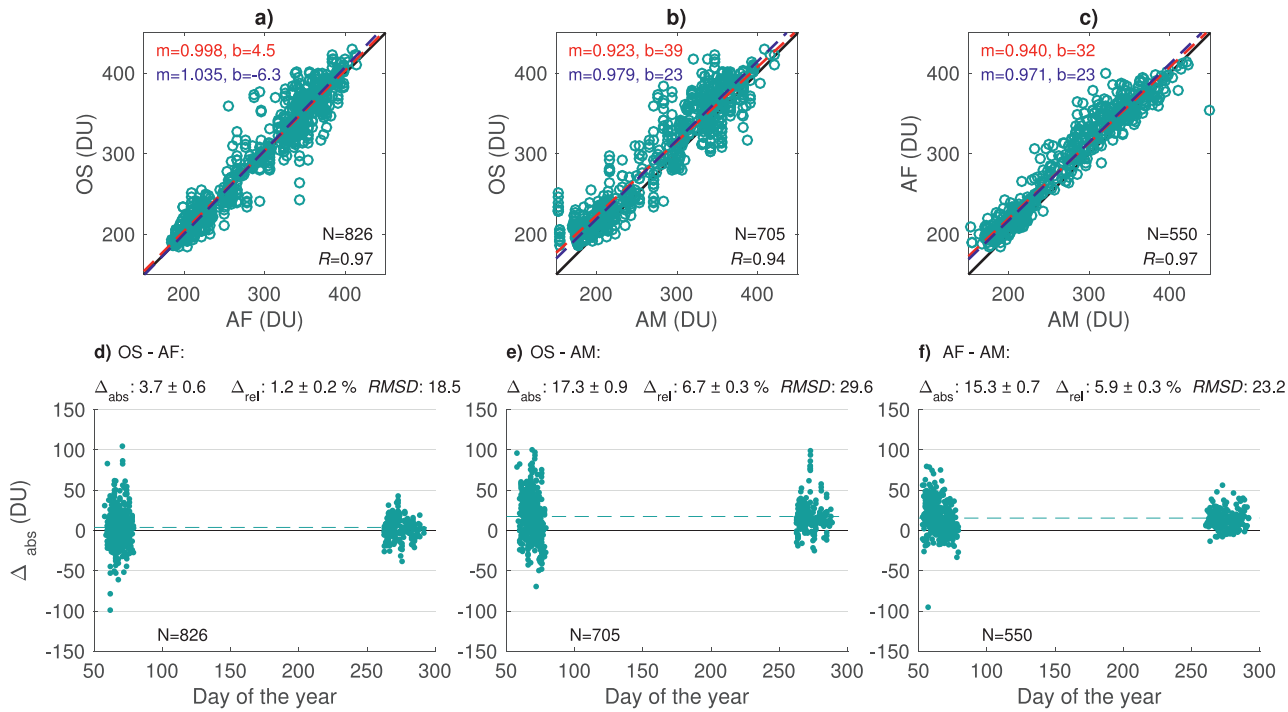

Fig. 1. Correlation plots (a–c) and seasonal absolute differences (d-f) between OSIRIS, ACE-FTS, and ACE-MAESTRO 14–52 km ozone partial columns. The correlation plots include best fit lines using the OLS (red dashed line) and RMA (blue dashed line) methods, as well as the one-to-one line (black). The slope, intercept, number of coincidences, and correlation coefficient are given as m, b, N, and *R*, respectively. In the difference plots, the dashed lines show the mean absolute difference. The errors shown for the mean differences and the *RMSD* values are the standard error. Abbreviations and measurement periods are given in Table 1.

FTIR and PARIS-IR averaging kernels, respectively. The magnitude of the change is similar for all satellite datasets. Smoothing with the ZSL-DOAS averaging kernels changes the spring and fall ozone columns by less than 1%. Satellite NO_2 partial columns change by less than 2% when smoothed with the Bruker FTIR averaging kernels. The change is less than 2.5% when smoothed with the ZSL-DOAS averaging kernels for the visible range, while smoothing with the UV averaging kernels leads to changes of 3–4%. All of the changes are small compared to the level of agreement between, and the combined error budgets of, the satellite minus ground-based instrument pairs for both ozone and NO_2.

Given the potential problems with the ZSL-DOAS averaging kernels, and the lack of Brewer averaging kernels, we preferred to treat all datasets in a consistent manner, and so we did not perform any smoothing for the satellite to ground-based comparisons.

5. Ozone results

5.1. Satellite versus satellite partial columns

Results of the comparisons between OSIRIS, ACE-FTS and ACE-MAESTRO 14–52 km ozone partial columns are shown in Fig. 1. The three datasets show good correlation, with correlation coefficients of 0.94 or greater (Fig. 1a–c). The slopes of the linear fits are close to 1, and the OLS and RMA methods agree well. The RMA fit is perhaps a better reference in this case, since none of the satellite datasets could be considered the reference dataset for the OLS fit. Correlation coefficients with the unknown truth (R^t from TCA) are 0.97 or greater for all three satellite instruments. Absolute differences between the satellite datasets are shown in Fig. 1d–f. OSIRIS and ACE-FTS show a mean relative difference of 1.2%. ACE-MAESTRO is systematically lower than OSIRIS and ACE-FTS, by 6.7% and 5.9%, respectively. The spread of the absolute differences (indicated by the *RMSD* value) is lowest for the OSIRIS minus ACE-FTS comparison, at 18.5 DU. *RMSD* values for ACE-MAESTRO are higher, 29.6 DU and 23.2 DU, when compared to OSIRIS and ACE-FTS, respectively. The *RMSD* values for the OSIRIS comparisons are within the maximum range expected from the *RMSE* calculations using TCA (Table 3), while the *RMSD* between the ACE instruments is outside the maximum expected range. The estimated values of the drift are 1.3 ± 2.4 %/decade for OSIRIS minus ACE-FTS, -2.1 ± 3.8%/decade for OSIRIS minus ACE-MAESTRO, and -2.1 ± 3.3%/decade for ACE-FTS minus ACE-MAESTRO. None of these values are statistically significant, indicating that there are no systematic changes between satellite datasets over time.

Previous versions of the ozone products from the three satellite instruments have been compared before. Fraser et al. [9] compared ACE-FTS v2.2 and ACE-MAESTRO v1.2 partial columns between 15 and 40 km in a 500 km radius around PEARL for 2004–2006, and found mean relative differences of 5.5% to 22.5%. The 2003–2017 mean of 5.9% found in this study falls within this range. Dupuy et al. [3] compared OSIRIS v2.1, ACE-FTS v2.2, and ACE-MAESTRO v1.2 profiles on a global scale for 2004–2006. They found that ACE-MAESTRO agreed with OSIRIS to ±7% in the 18–59 km range, while ACE-FTS was on average 6% larger than OSIRIS between 9 and 45 km, and progressively larger (up to 44%) between 45 and 60 km. This is opposite to the findings of this study, where ACE-FTS and ACE-MAESTRO partial columns are both smaller than OSIRIS partial columns. The discrepancy is likely due to the fact that coincidences in this study are limited to the Arctic, while Dupuy et al. [3] covered all latitudes. This conclusion is also supported by Adams et al. [8], who compared OSIRIS v5.0x, ACE-FTS v3.0, and ACE-MAESTRO v1.2 partial columns for 14–52 km (same altitude range as in this study) near PEARL for 2004–2010. Mean relative differences between OSIRIS and ACE-FTS were reported to be 1.2%, identical to the value found in this study. Comparisons involving ACE-MAESTRO partial columns show an approximate doubling

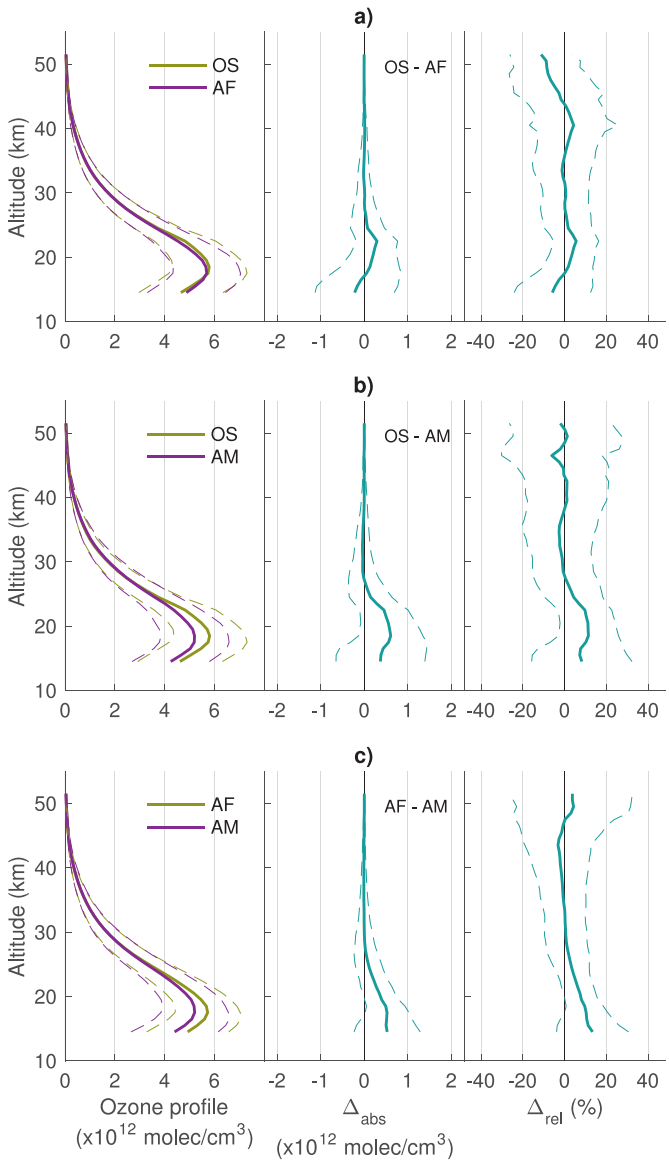

Fig. 2. Mean ozone number density profiles and mean differences for all coincidences between OSIRIS, ACE-FTS and ACE-MAESTRO. The left panels show the mean profiles, with one standard deviation limits indicated by the dashed lines. The middle and right panels show the absolute and relative differences, respectively, at each altitude level.

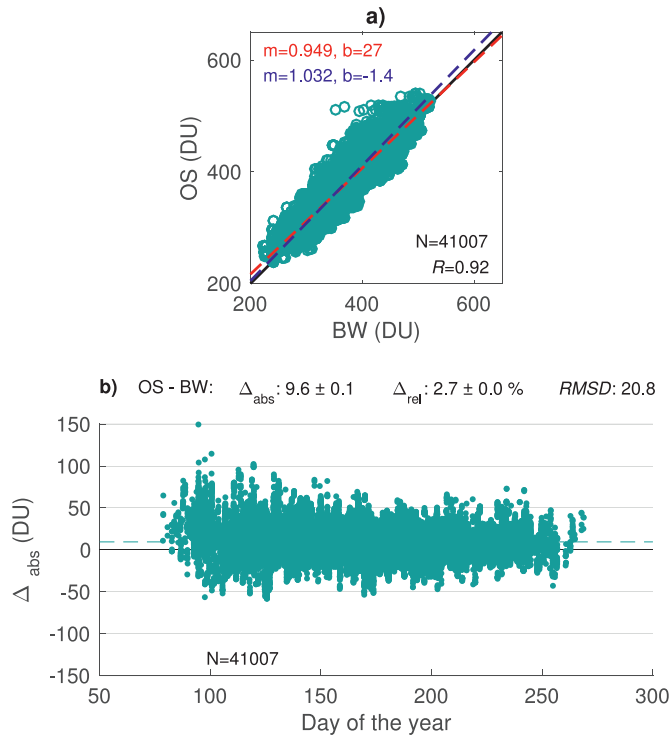

Fig. 3. As for Fig. 1, OSIRIS-plus-sonde surface-52 km ozone columns and Brewer total columns.

of the relative differences, from 2.8% [8] to 6.7% and 5.9%. Given that the OSIRIS minus ACE-FTS comparison remained unchanged, this difference is likely due to changes in the more recent ACE-MAESTRO v3.13 dataset. The relative differences show the same doubling for the 2004–2010 period (used by Adams et al. [8]), indicating that the issue is related to the v3.13 processing. Adams et al. [8] also reported slopes significantly less than 1 for OSIRIS minus ACE-MAESTRO and ACE-FTS minus ACE-MAESTRO comparisons, with y-intercepts similar to those shown in red in Fig. 1b,c.

To further investigate this apparent low bias in ACE-MAESTRO data, we compared 14–52 km ozone number density profiles for all three satellite instruments. The mean profiles and standard deviations for all coincidences are shown in Fig. 2. ACE-MAESTRO underestimates the peak ozone concentrations compared to both OSIRIS (Fig. 2b) and ACE-FTS (Fig. 2c), by more than 10%. OSIRIS and ACE-FTS profiles agree well (Fig. 2a), with only a small difference in the altitude of the peak ozone concentrations. The agreement above 25 km is good for all instrument pairs. ACE-FTS number densities are larger than OSIRIS above 45 km, consistent with Dupuy et al. [3].

5.2. Satellite versus ground-based partial columns

Correlation plots of the satellite-plus-sonde ozone columns (surface-52 km) and the ground-based datasets are shown in Figs. 3 and 4. Comparisons with the Brewer ozone data (Fig. 3) are only shown for OSIRIS, since there are too few (less than 15) Brewer measurements in early spring and late fall for meaningful comparisons with ACE. The instrument pairs have correlation coefficients of 0.86-0.95 for OSIRIS, 0.90-0.96 for ACE-FTS, and 0.87-0.94 for ACE-MAESTRO. The ZSL-DOAS instruments show better correlation with the ACE instruments than the direct sun measurements, while OSIRIS shows high correlation coefficients for all instruments except PARIS-IR. R^t values from TCA are 0.94-0.97 for OSIRIS, 0.94-0.96 for ACE-FTS, and 0.92-0.94 for ACE-MAESTRO. R^t for the ground-based instruments ranges from 0.92 to 0.98. The seasonal evolution of the absolute differences between the instrument pairs, as well as the mean absolute and relative differences and *RMSD* values for each pair are shown in Fig. 5. Most instrument pairs (with the exception of OSIRIS minus Brewer, ACE minus Bruker FTIR, and ACE-MAESTRO minus PARIS-IR) agree within the combined retrieval uncertainties (absolute and relative) indicated in Table 2. Note that the error estimates for the satellite data and for the Brewer measurements include random errors only.

The comparison of OSIRIS-plus-sonde ozone columns to Brewer data shows a mean relative difference of 2.7%, with the largest differences observed in the spring (Fig. 3). The vast majority of the coincidences, however, occur in the summer, and so the larger springtime differences contribute minimally to the mean. The relative differences (not shown) are distributed evenly throughout the year. For a discussion of the dependence of the differences on SZA,

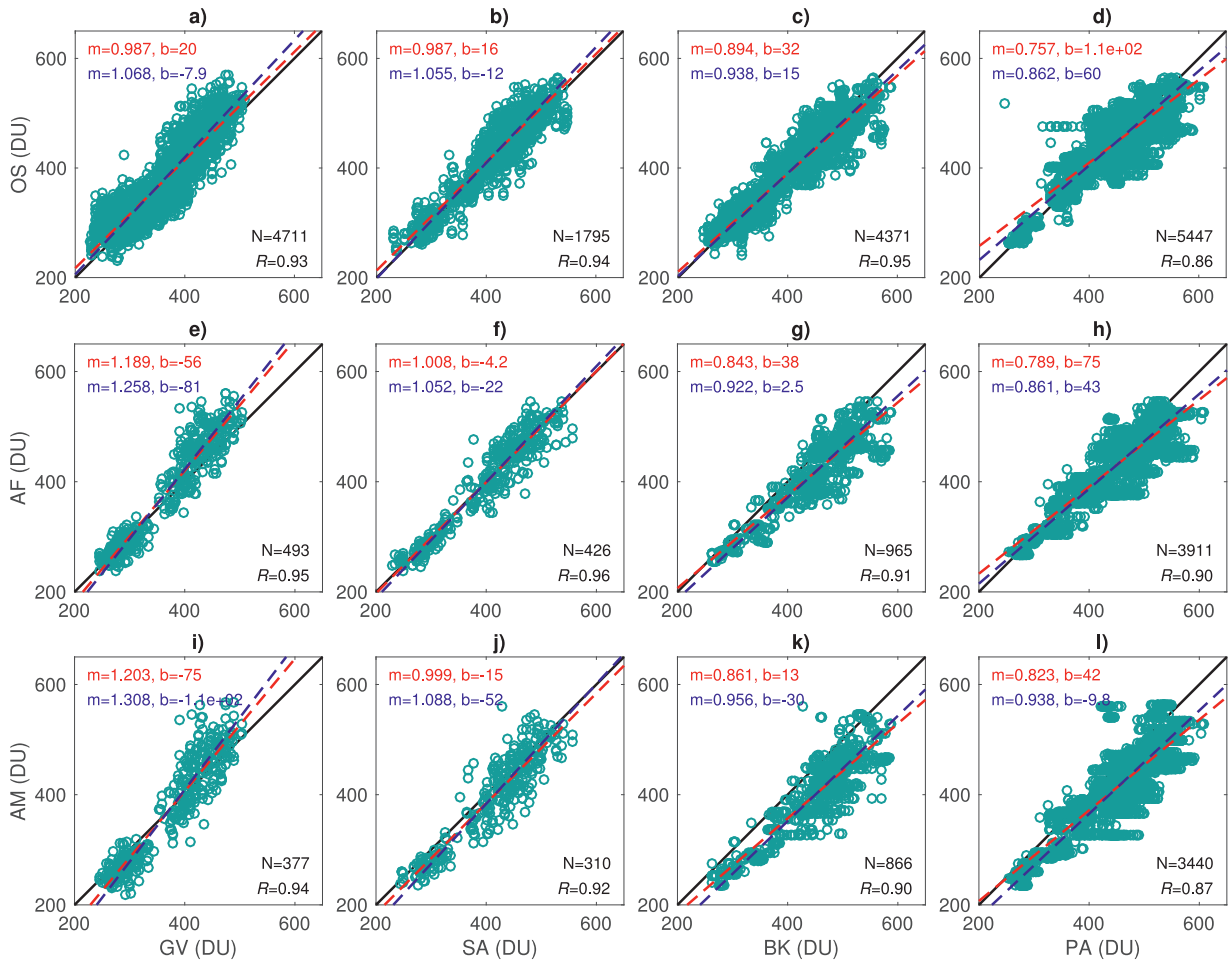

Fig. 4. Correlation plots for satellite-plus-sonde surface-52 km ozone columns (*y*-axes) against the ground-based total columns (*x*-axes). The plots include best fit lines using the OLS (red dashed line) and RMA (blue dashed line) methods, as well as the one-to-one line (black). The slope, intercept, number of coincidences, and correlation coefficient are given as m, b, N, and *R*, respectively. Abbreviations and measurement periods are given in Table 1.

see Appendix A. The *RMSD* value of 20.8 DU is within the expected range from the *RMSE* calculations shown in Table 3.

OSIRIS and ACE-FTS satellite-plus-sonde columns are consistently larger than the GBS ozone columns, by 4.4% and 2.6%, respectively. The absolute differences are most pronounced for the higher ozone values in early spring. OSIRIS and ACE-FTS show better agreement with the SAOZ dataset across the range of ozone column values, with mean relative differences of 2.3% and -0.5%, respectively. ACE-MAESTRO ozone is systematically lower than OSIRIS and ACE-FTS, and therefore agrees better with GBS (-1.2%) than SAOZ (-4.4%). The offset between the GBS and SAOZ intercomparisons is similar for both ACE instruments. The largest absolute differences (as well as relative differences, not shown) for each satellite minus ZSL-DOAS instrument pair are observed in the early spring (Fig. 5a, c, e). The *RMSD* values for the satellite minus ZSL-DOAS comparisons are all within the maximum expected range shown in Table 3. Comparisons to the GBS dataset consistently result in higher *RMSD* (30.4 DU, 33.9 DU, and 36.5 DU for OSIRIS, ACE-FTS, and ACE-MAESTRO) than comparisons to SAOZ (26.7 DU, 24.4 DU, and 35.5 DU, respectively). This difference is smallest for ACE-MAESTRO, and the highest *RMSD* values are also seen in the ACE-MAESTRO comparisons. To aid in interpreting the intercomparison results, the dependence of the differences on SZA is described in Appendix A, and the ground-based ozone datasets are compared in Appendix B.1.

All three satellite-plus-sonde ozone datasets are systematically lower than the Bruker FTIR. This difference (absolute and relative) is also most pronounced in the spring, resulting in large mean relative differences for ACE-FTS and ACE-MAESTRO comparisons, -7.5% and -12.0%, respectively. In the case of OSIRIS, the agreement is -2.1%, and it remains better than 3% in all seasons. The satellite-plus-sonde columns show better agreement with the PARIS-IR, resulting in mean relative differences of -4.3% for ACE-FTS, -8.8% for ACE-MAESTRO, and -0.1% for OSIRIS.

Comparisons of 14–52 km satellite partial columns to Bruker FTIR and PARIS-IR partial columns show small changes in relative differences (compared to surface-52 km satellite-plus-sonde columns) for the Bruker FTIR, to -2.3%, -7.3%, and -13.4% for OSIRIS, ACE-FTS and ACE-MAESTRO. These changes are significant within standard error for ACE-MAESTRO only. PARIS-IR differences show larger (and significant) changes, to 1.5%, -0.4%, and -6.1%, respectively. Results using PARIS-IR partial columns, however, need to be interpreted with caution, since the retrieval is optimized for total columns, and has lower DOFS than the Bruker FTIR retrieval.

The *RMSD* values (using surface-52 km satellite-plus-sonde ozone columns) are 25.2 DU, 45.1 DU, and 60.7 DU for OSIRIS, ACE-FTS, and ACE-MAESTRO, when compared to the Bruker FTIR. The values are 33.3 DU, 37.8 DU, and 53.8 DU, respectively, when compared to PARIS-IR. For the Bruker FTIR, only the OSIRIS comparison falls in the expected range from the *RMSE* values (Table 3),

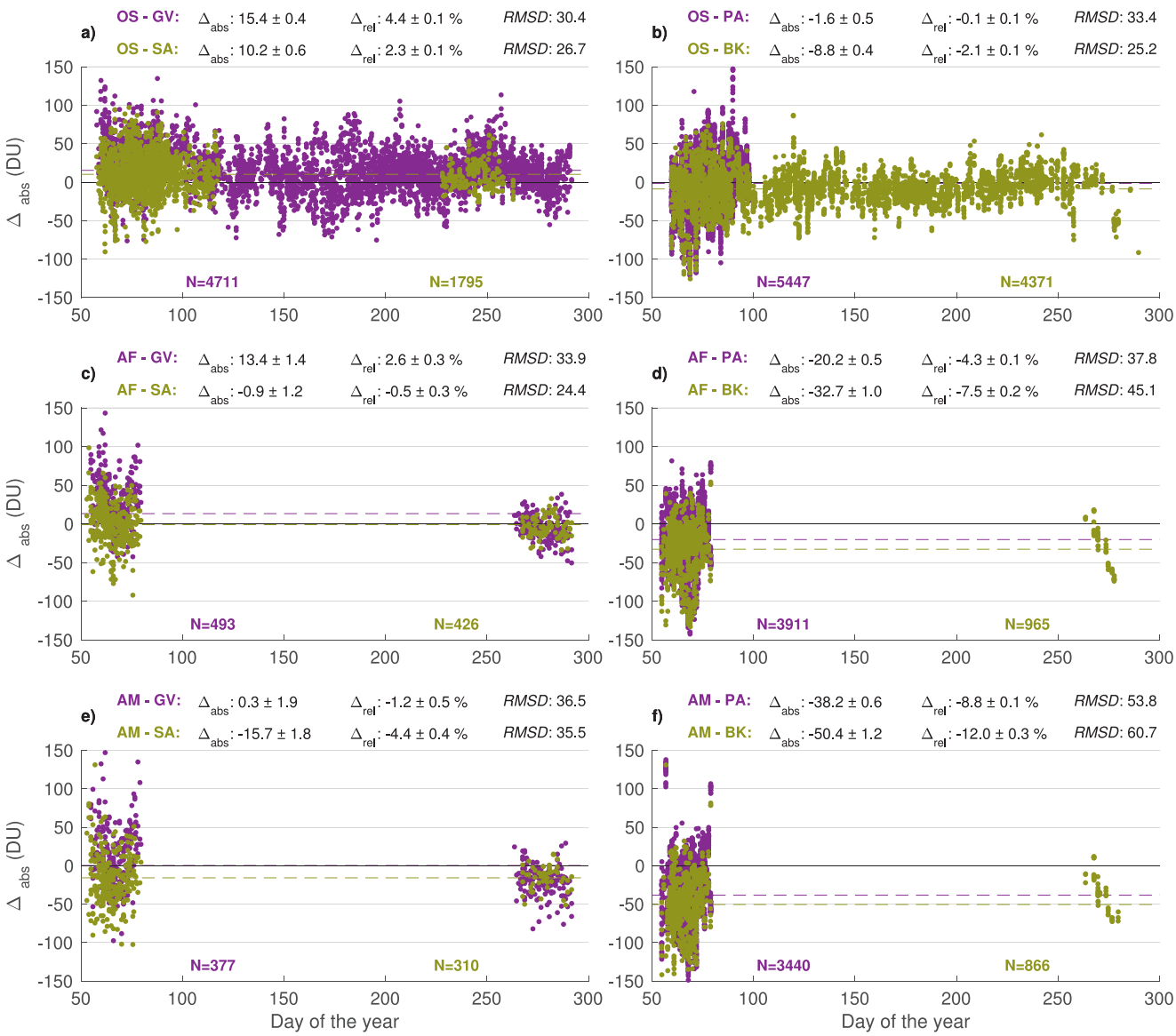

Fig. 5. Seasonal absolute differences between satellite-plus-sonde surface-52 km ozone columns and the ground-based datasets. The dashed lines represent the mean absolute differences. The errors shown for the mean differences and the *RMSD* values are the standard error. Abbreviations and measurement periods are given in Table 1.

while PARIS-IR satisfies the *RMSE* condition for OSIRIS and ACE-FTS.

Comparisons of 14–52 km ozone profiles from the Bruker FTIR and the satellite instruments (linearly interpolated to the Bruker FTIR retrieval grid) are shown in Fig. 6. PARIS-IR profiles were not used due to the comparatively low DOFS of the PARIS-IR retrievals. OSIRIS profiles show good agreement with the Bruker FTIR profiles; the mean values are within 5% for all but the lowermost three layers. ACE-FTS and ACE-MAESTRO show patterns similar to each other, with the ACE-MAESTRO differences shifted due to the systematic differences discussed in Section 5.1. The ACE-FTS and ACE-MAESTRO profiles below 40 km are smaller than the Bruker FTIR values by as much as 12% and 20%, respectively, while relative differences above 40 km are of similar magnitude but with opposite sign. The large differences in the ACE minus Bruker FTIR column intercomparisons are the result of the large differences in the measured peak ozone concentrations. When only early spring data are considered for OSIRIS, the relative differences show a pattern similar to the ACE instruments, but with less of a difference below 40 km. The high-altitude differences may be related to the fast-decreasing vertical resolution of the Bruker FTIR above 30 km. To test if the discrepancies are due to the different vertical resolutions of the satellite instruments and the Bruker FTIR, the profile comparisons were repeated using satellite profiles smoothed with the Bruker FTIR averaging kernels. The new comparisons (not shown) are similar to the unsmoothed results, indicating that smoothing does not have a large impact on the mean profile comparisons. The springtime measurements are likely affected by the location of the polar vortex; this is examined in Section 7.1.

5.3. Comparison to previous validation studies

The ZSL-DOAS instruments at Eureka have been used in several satellite validation studies. Fraser et al. [9] compared ACE-FTS v2.2 and ACE-MAESTRO v1.2 15–40 km ozone partial columns (extended with ozonesonde data) to 2004–2006 GBS and SAOZ columns. The GBS and SAOZ ozone was retrieved using identical settings in that study. When comparing ACE-FTS to ZSL-DOAS data, they found

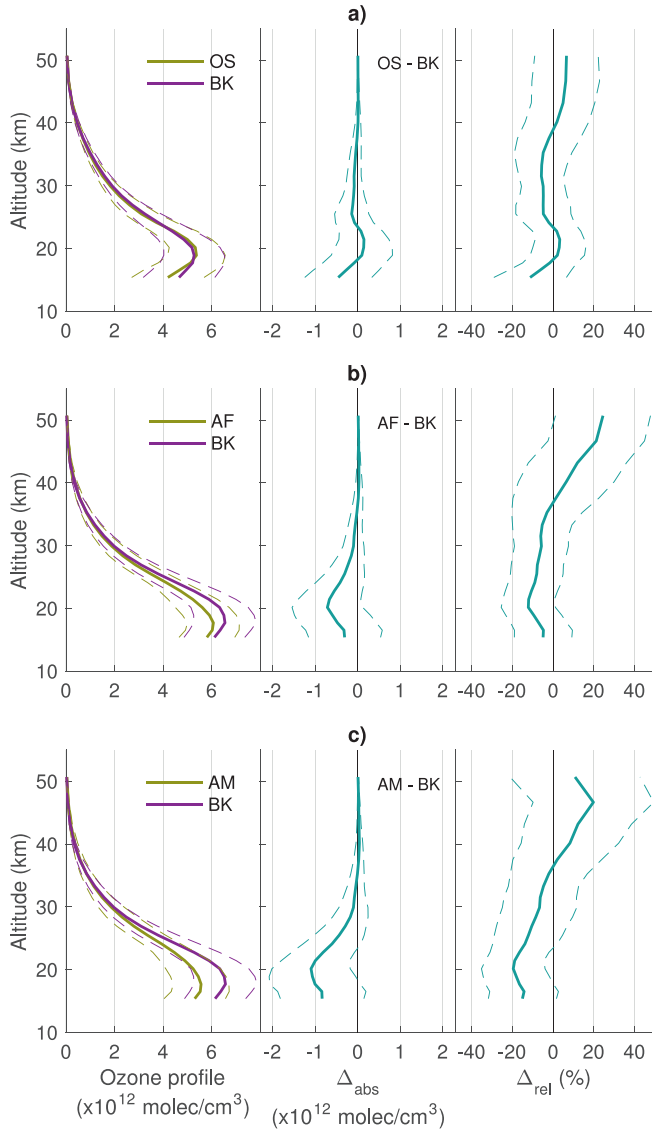

Fig. 6. As for Fig. 2, satellite profiles against Bruker FTIR profiles.

mean relative differences of 3.2% to 6.3% for GBS ozone, and 0.1% to 4.3% for SAOZ. These values are comparable to the 2.6% and -0.5% found in this study. For ACE-MAESTRO, Fraser et al. [9] found differences of −19.4% to −1.2% for GBS and −12.9% to −1.9% for SAOZ. Our values of −1.2% and −4.4% are within the range estimated by Fraser et al. [9]. Adams et al. [8] compared OSIRIS v5.0x, ACE-FTS v3.0 and ACE-MAESTRO v1.2 ozone columns with GBS and SAOZ V2 measurements for 2003–2011 using methodology similar to the methods in this paper. For OSIRIS, they found differences of 5.7% and 7.3% with respect to GBS and SAOZ data, which are larger than the 4.4% and 2.3% reported in this study. Since the present study also uses the OSIRIS v5.x data, the reduction in the differences with respect to the GBS measurements is largely due to the longer data record, while the SAOZ intercomparisons were improved by the new SAOZ V3 dataset as well (Section 3.1; V3 ozone is significantly larger than V2 data for 2008–2010). Adams et al. [8] reported ACE-FTS relative differences of 6.5% and 4.8% for GBS and SAOZ, which are also larger than the 2.6% and -0.5% found in this study. In addition to the reasons mentioned before, this improvement is largely due to the addition of more fall ACE-FTS data, which generally agrees better with the ZSL-DOAS datasets (Fig. 5c). ACE-MAESTRO relative differences changed from 5.0% and 1.6% [8] to −1.2% and −4.4% for GBS and SAOZ, respectively, reflecting the apparent negative bias in the new ACE-MAESTRO dataset.

Bruker FTIR ozone was first compared to ACE-FTS v2.2 measurements by Batchelor et al. [10]. They compared 6–43 km partial columns to ACE-FTS partial columns smoothed by the Bruker FTIR averaging kernels, and found a mean relative difference of -5.6% for 2007–2008. This is comparable to the -7.5% relative difference reported in this study. Batchelor et al. [10] found that the location of the polar vortex had a significant impact on the comparison results. Implementing stricter coincidence criteria based on line-of-sight scaled potential vorticity (sPV) and temperature values improved the relative differences to -0.4%. The impact of the vortex position in the results of this study is further discussed in Section 7.1. Using the stricter coincidence criteria of Batchelor et al. [10], Griffin et al. [12] compared smoothed ACE-FTS v3.5 ozone partial columns to Bruker FTIR partial columns in the 9–48.5 km range. They found mean relative differences of -3.6% for 2007–2013, smaller than the value found in this study. PARIS-IR ozone has only been compared to ACE-FTS previously. Fu et al. [11] compared 2006 measurements to smoothed ACE-FTS v2.2 data in the 9.5–84.5 km range, and found a mean relative difference of −5.2%, while Griffin et al. [12] found −3.5%. Both these values are similar to the −4.3% reported here.

Adams et al. [8] compared Bruker FTIR total columns to OSIRIS v5.0x, ACE-FTS v3.0 and ACE-MAESTRO v1.2 satellite-plus-sonde columns using methods similar to the ones applied here, and found mean relative differences of 0.1%, −4.7%, and −6.1%, respectively. These values are smaller than the values of −2.1%, −7.5%, and −12.0% found in this study. Most of the differences can be explained by year-to-year variability introduced by the polar vortex in the spring (see Section 7.1), and by the shift in the ACE-MAESTRO data. Adams et al. [8] also compared Brewer ozone total columns to OSIRIS-plus-sonde columns, and found a mean relative difference of 2.8 %, very close to the 2.7% in this study. The two values agree within their combined standard errors. The comparison results for satellite and ground-based ozone columns from this study and from relevant publications are summarized in Fig. 9a.

5.4. Decadal stability

Drift values and corresponding uncertainties for each of the relative difference time series are shown in Table 4. OSIRIS-plus-sonde ozone columns show a statistically significant drift of 2.7 ± 1.3%/decade only when compared to the Brewer measurements. The mean drift also becomes significant as a result. The number of years required to detect a real drift of 2.7 %/decade (see Section 4.4), however, is n^*=23, while the OSIRIS to Brewer comparisons span only 14 years. In addition, OSIRIS shows no significant drift when compared to any other ground-based dataset, and so we cannot say with confidence that the drift between OSIRIS-plus-sonde ozone columns and Brewer measurements is real. Hubert et al. [51] found significant drifts in the differences between OSIRIS ozone data and ozonesonde and lidar measurements. These issues, however, were related to a pointing bias, and were corrected in the v5.10 dataset [23] (see Section 2.7).

ACE-FTS-plus-sonde ozone columns show no statistically significant drift when compared to any of the ground-based instruments, and the n^* values indicate that none of the time series are long enough to say with confidence that the drifts returned by the linear regression are real. When the mean across all instrument pairs is considered, the drift becomes significant, since the combined uncertainty is reduced. This apparent negative drift is expected, given the better agreement of fall ACE-FTS data with ZSL-DOAS measurements, and the fact that most fall coincidences occur after 2013 (see Section 5.3). The mean drift is not significant when

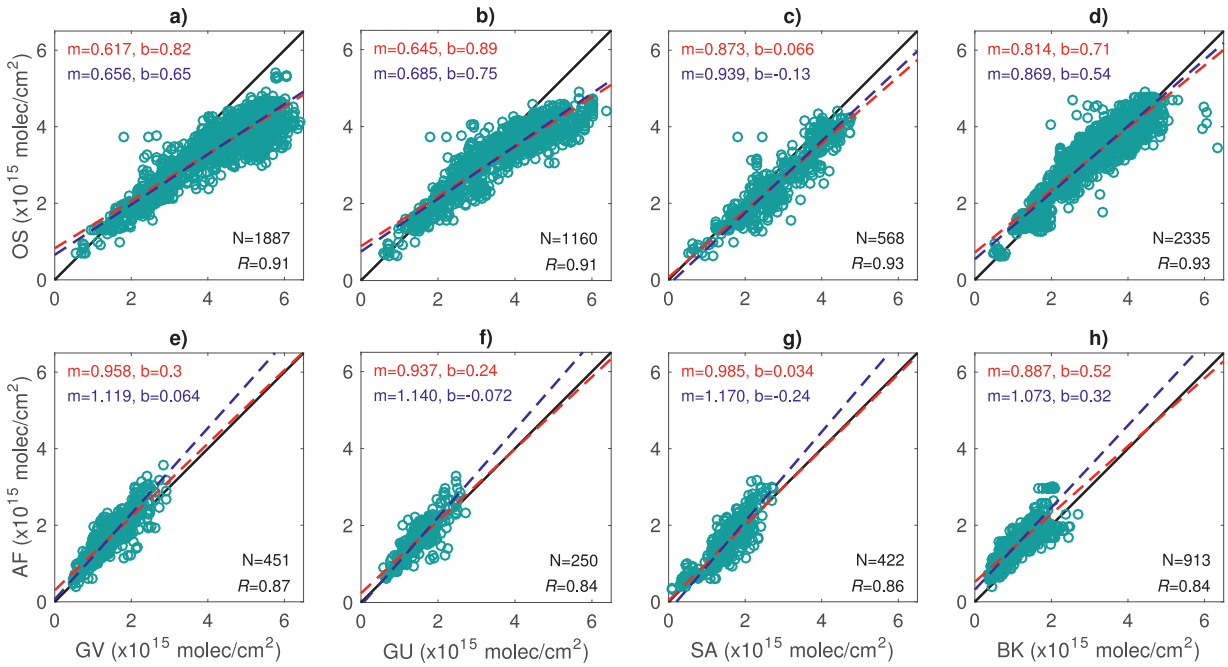

Fig. 7. As for Fig. 4, 12–40 km NO_2 satellite partial columns against ground-based partial columns.

fall ZSL-DOAS data are excluded, or when the spring and fall data are fitted separately. The lack of drift in ACE-FTS data is consistent with the results of Hubert et al. [51].

There is no significant drift between ACE-MAESTRO-plus-sonde ozone columns and the individual ground-based instruments, and the mean drift is also zero within the uncertainty, in agreement with Hubert et al. [51]. All of the n^* values are larger than the number of years available in each relative difference time series. The uncertainties on the drift values are larger than for the ACE-FTS data, reflecting the larger scatter seen in the comparison results (Fig. 5). The lack of drift in the ACE-MAESTRO data lends further credibility to the conclusion that the observed low bias (Section 5.1) is related to the v3.13 reprocessing, and not to changes in the dataset over time.

6. NO_2 results

Satellite NO_2 measurements were only compared to the ground-based datasets. The comparison of OSIRIS and ACE-FTS NO_2 was excluded due to the limited number (38) and seasonal distribution (late September in a few years only) of coincident measurements. The results of the satellite minus ground-based intercomparisons are discussed below. Unlike for ozone, NO_2 profiles from the Bruker FTIR were not compared to the satellite profiles, since the mean degrees of freedom for signal for the Bruker FTIR 12–40 km partial columns is 1.2 on average.

6.1. Satellite versus ground-based partial columns

Correlation plots of the satellite and ground-based 12–40 km NO_2 partial columns are shown in Fig. 7. Correlation coefficients for OSIRIS are in the 0.91–0.93 range. The values are slightly smaller for ACE-FTS, between 0.84–0.87. One reason for this might be that ACE-FTS only measures in the spring and fall, and so only a smaller range of NO_2 partial column values is available to constrain the linear relationship. R^t coefficients from TCA are 0.94–0.96 for OSIRIS and 0.88–0.92 for ACE-FTS, while the ground-based datasets have R^t values in the 0.88–0.97 range. The absolute differences between the instrument pairs throughout the year are shown in Fig. 8, alongside the mean absolute and relative differences and *RMSD* values. Most instrument pairs (with the exception of OSIRIS minus GBS-vis and ACE-FTS minus Bruker FTIR) agree within the combined retrieval uncertainties (absolute and relative) indicated in Table 2.

OSIRIS NO_2 shows a similar relationship to the GBS-vis and GBS-UV products (Fig. 8a), where there is good agreement in the spring, OSIRIS partial columns are much smaller than the ground-based data in the summer, and this difference is reduced in the fall. The relative differences (not shown) follow the same pattern. The mean relative differences are −19.9% and −8.1% with respect to GBS-vis and GBS-UV, respectively. Since the GBS-vis dataset is longer and has more summer data, the mean differences are heavily weighted by the large differences in the summer. OSIRIS NO_2 is also smaller than SAOZ, with a relative difference of −11.3% (Fig. 8b), and similar absolute differences in the spring and fall. The relative differences (not shown) are larger in the spring for the SAOZ comparisons. The *RMSD*s for the GBS datasets (10.3–6.6 $\times 10^{14}$ molec/cm^2) are larger than the maximum expected spread from the *RMSE* calculations (Table 3), likely due to the large summertime differences. The *RMSD* for SAOZ is smaller (4.8 $\times 10^{14}$ molec/cm^2), and within the expected range. The $SAOZ_{allyear}$ dataset (Section 3.1) provides four years of summer data to further evaluate the differences between OSIRIS and ground-based ZSL-DOAS instruments. The OSIRIS minus $SAOZ_{allyear}$ NO_2 comparison (not shown) results in a mean relative difference of −10.5%, and the differences follow the same pattern described for the GBS instruments. The large summertime OSIRIS minus ZSL-DOAS differences suggest the presence of systematic errors in the datasets or in the scaling factors from the photochemical box model [8]. The differences in viewing geometries, combined with the challenges of ZSL-DOAS retrievals in the summer (Section 3.1) likely contribute to the seasonal pattern as well.

OSIRIS NO_2 partial columns are larger than Bruker FTIR measurements by a mean difference of 5.5%. The seasonal pattern in the absolute differences (Fig. 8b) and relative differences (not shown) is somewhat similar to the ZSL-DOAS datasets, although

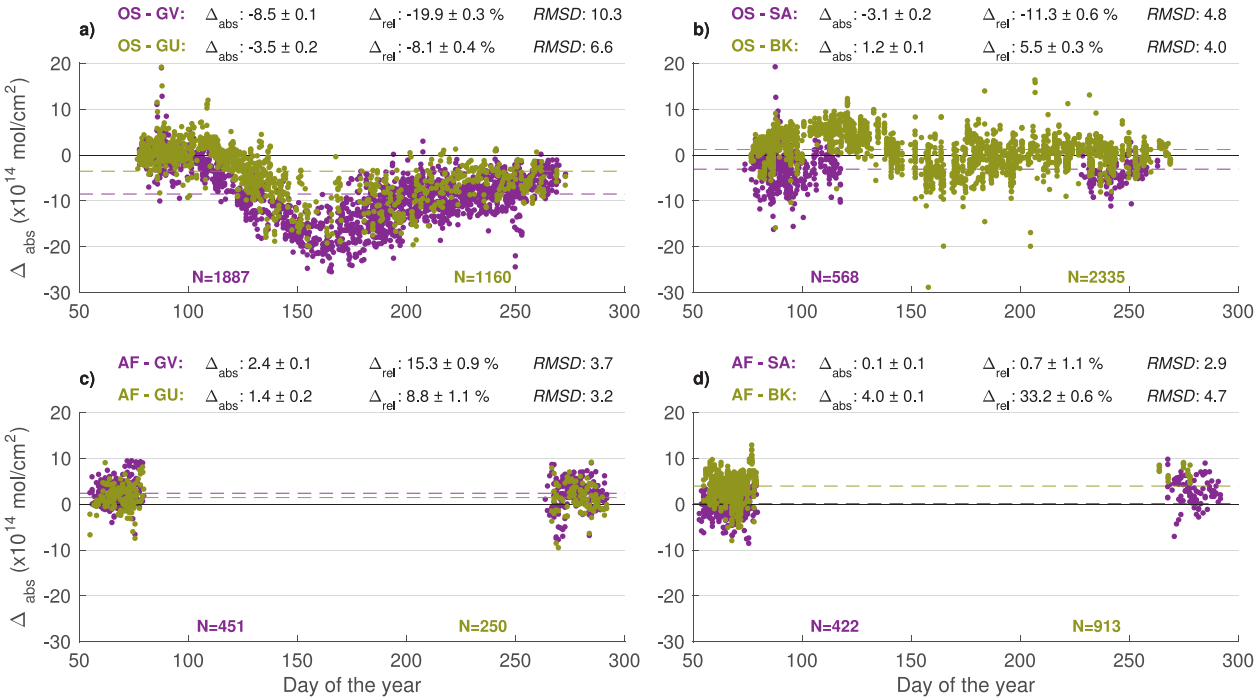

Fig. 8. As for Fig. 5, 12–40 km NO_2 satellite partial columns against ground-based partial columns.

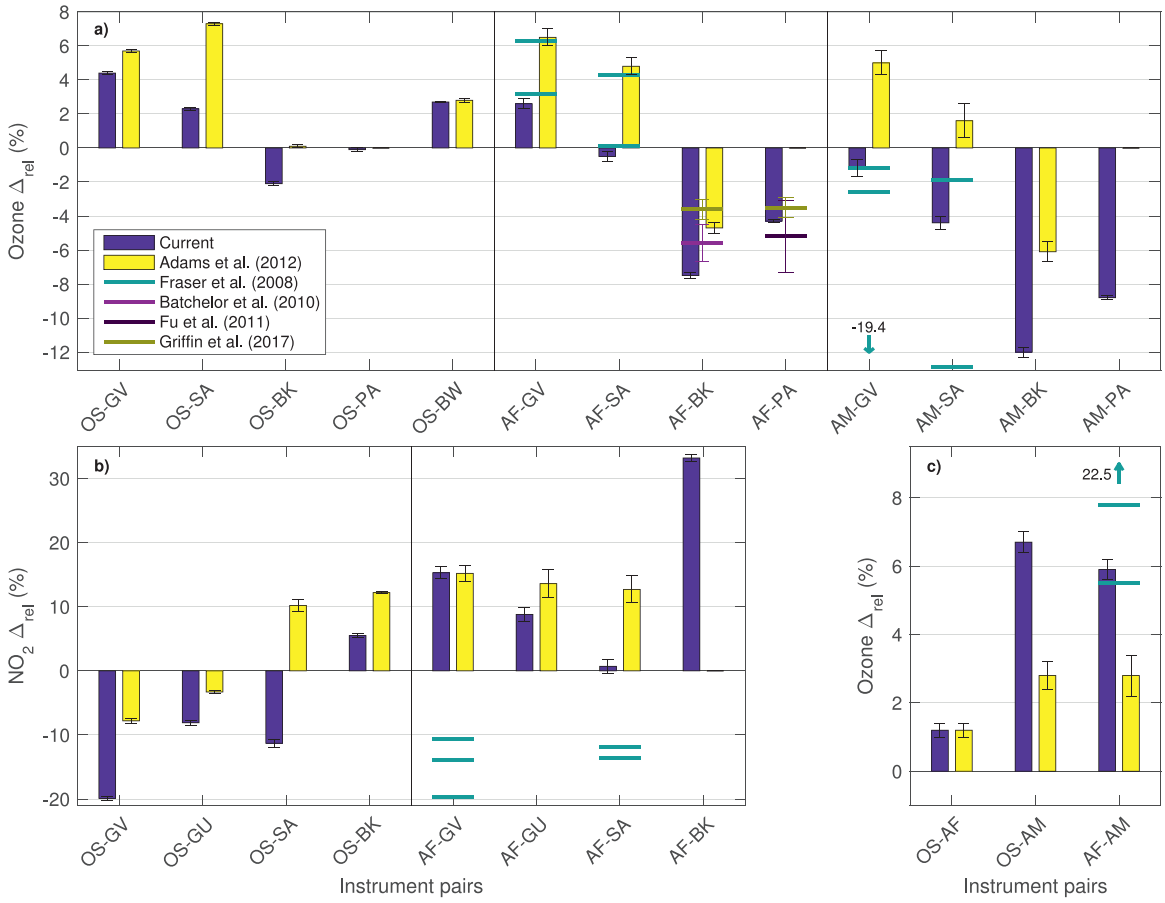

Fig. 9. Mean relative differences between (a) satellite-plus-sonde surface-52 km ozone columns and ground-based total columns, (b) 12–40 km satellite NO_2 partial columns and ground-based partial columns, and (c) 14–52 km satellite ozone columns. Bars show the results from this study, as well as the results from Adams et al. [8] (2003–2011). Additional lines indicate results from Fraser et al. [9] (individual yearly values, 2004, 2005, and 2006 for GBS, 2005 and 2006 for SAOZ; note that for AF-GV, values are 6.3% for both 2005 and 2006), Batchelor et al. [10] (2007–2008), Fu et al. [11] (2006), and Griffin et al. [12] (2006–2013). Error bars indicate standard error, where available. Abbreviations are given in Table 1.

the summertime negative shift in the absolute differences is less pronounced. Overall, OSIRIS shows better agreement with Bruker FTIR than with the ZSL-DOAS instruments. This is reflected in the *RMSD* (4.0×10^{14} molec/cm^2), which is the smallest among the four OSIRIS comparisons. The differences between the ground-based NO_2 datasets are not unusual, and are discussed further in Appendix B.2. The dependence of the differences on SZA is described in Appendix A.

The ACE-FTS NO_2 partial columns are systematically larger than the GBS-vis and GBS-UV datasets, with mean differences of 15.3% and 8.8%, respectively (Fig. 8c). The SAOZ data, on the other hand, agree well with ACE-FTS, with a mean difference of only 0.7% (Fig. 8d). The level of agreement is similar in spring and fall for the GBS datasets, while SAOZ appears to measure less NO_2 in the fall compared to ACE-FTS. The ACE-FTS NO_2 partial columns are also smaller than the Bruker FTIR, by 33.2% on average. The majority of coincidences occur in the spring, due to the limited number of Bruker FTIR measurements in the fall. The *RMSD* for the GBS-vis and GBS-UV datasets (3.7–3.2 $\times 10^{14}$ molec/cm^2) are comparable, but the GBS-vis value falls outside the range indicated in Table 3. The ACE-FTS minus SAOZ comparison shows the smallest spread (2.9×10^{14} molec/cm^2, within the expected range), while the Bruker FTIR shows the largest (4.7×10^{14} molec/cm^2, outside the expected range).

6.2. Comparison to previous validation studies

Fraser et al. [9] compared ACE-FTS v2.2 22–40 km NO_2 partial columns to 2004–2006 GBS-vis and SAOZ columns. They found mean relative differences of −10.7% to −19.7% for GBS-vis, and −11.9% to −13.6% for SAOZ. These values are opposite of the findings in this study, but direct comparisons are difficult due to the different partial column range and the fact that Fraser et al. [9] calculated total columns (instead of the 12–60 km range used here) for the ground-based instruments. More direct comparison is possible to the results of Adams et al. [8], who used settings similar to the ones in this study for the GBS datasets. They calculated 17–40 km partial columns for the satellite intercomparisons, using the OSIRIS v3.0 and ACE-FTS v2.2 datasets. For OSIRIS, they found mean relative differences of −7.8% and −3.3% for GBS-vis and GBS-UV. The differences with respect to the GBS datasets are much larger (−19.9% and −8.1%) in this study. This is primarily the result of the different seasonal distribution of the OSIRIS v6.0 NO_2 measurements. Near PEARL, there are fewer OSIRIS measurement in the spring and fall, when the agreement with the ground-based partial columns is better. Adams et al. [8] also found a seasonal variation in the absolute differences similar to what is shown in Fig. 8a. For ACE-FTS, they found relative differences of 15.2% and 13.6% using GBS-vis and GBS-UV data, similar to the 15.3 % and 8.8% found in this study. The larger change in the GBS-UV comparison is likely due to the low number of coincidences (38) in Adams et al. [8]. Using SAOZ V2 data, they found mean relative differences of 10.2% for OSIRIS and 12.7% for ACE-FTS, which are substantially different from the values of −11.3% and 0.7% found in this study. However, changes in the SAOZ NO_2 AMF calculations (Section 3.1) likely account for most of these differences. The significant changes in the SAOZ dataset are evident in the ground-based intercomparisons as well (see Appendix B.2).

The Bruker FTIR NO_2 has only been used in one previous validation study. Using a previous version of the retrievals, Adams et al. [8] compared 17–40 km OSIRIS partial columns to Bruker FTIR partial columns. They found a mean relative difference of 12.2% for 2006–2011, and the differences showed a seasonal pattern similar to the OSIRIS minus ZSL-DOAS comparisons. The mean OSIRIS minus Bruker FTIR difference is smaller in this study (5.5%), and the seasonal variation is less pronounced, since the updated Bruker FTIR NO_2 shows better agreement with OSIRIS for low NO_2 concentrations in the spring and fall. ACE-FTS NO_2 measurements have not previously been compared to the Bruker FTIR at Eureka, since Adams et al. [8] excluded Bruker FTIR measurements with SZA > 80°. The satellite to ground-based NO_2 partial column comparison results, and the changes compared to previous publications, are summarised in Fig. 9b.

6.3. Decadal stability

Drift values and corresponding uncertainties for each of the relative difference time series are shown in Table 4. There is no statistically significant drift between OSIRIS NO_2 partial columns and most ground-based instruments, except for a marginally significant drift of -10.4 ± 10.3%/decade when compared to the Bruker FTIR dataset. The mean drift for OSIRIS NO_2 partial columns is not statistically significant. The drift with respect to the Bruker FTIR dataset is related to the fact that springtime coincidences show large positive differences (but also large scatter), and the majority of springtime coincidences occur prior to 2012. This leads to a negative drift in the springtime comparisons. None of the other instruments show a significant drift for spring data only, and the OSIRIS minus Bruker FTIR drift is not significant when spring data are excluded. In addition, the number of years required to detect a drift of −10.4%/decade in the OSIRIS minus Bruker FTIR dataset is n^*=20, and there are only 12 years of coincident measurements available.

ACE-FTS NO_2 partial columns show no statistically significant drift when compared to any of the ground-based datasets. The mean drift, however, is statistically significant. Similar to the ozone comparisons, this apparent positive drift is expected, since ACE-FTS minus SAOZ differences are larger in the fall than in the spring (see Section 6.1), and fall coincidences only occur in 2015–2017. The mean drift is not significant when fall SAOZ data are excluded, or when the spring and fall data are fitted separately. In addition, n* values for each of the individual drifts are larger than the number of years available in the relative difference time series.

7. The impact of atmospheric conditions

7.1. Springtime coincidence criteria

Many of the instrument comparisons in this study show the poorest agreement during the spring, when the polar vortex might be located over or near Eureka. The vortex isolates airmasses in the stratosphere, and so measurements on either side of the vortex boundary might be spatially close, but have substantially different trace gas concentrations. To examine the effect of the polar vortex on the springtime comparisons, we used derived meteorological products (DMPs) [59] from the second Modern-Era Retrospective analysis for Research and Applications (MERRA-2), an atmospheric reanalysis that uses the Goddard Earth Observing System Model Version 5.2.0 reanalysis system (GEOS-5) [60,61]. DMPs, such as scaled potential vorticity (sPV) [62,63] and temperature, were calculated along the line-of-sight of the ACE-FTS, ACE-MAESTRO, GBS, SAOZ, Bruker FTIR and PARIS-IR instruments, and at the coordinates of the 25 km tangent point for OSIRIS measurements, using the Jet and Tropopause Products for Analysis and Characterization (JETPAC) package [64,65].

The line-of-sight calculations for the ZSL-DOAS instruments are described by Adams et al. [8]. Coincident measurements were kept only if the temperature differences between measurements at selected layers were less than 10 K, and measurements were either both inside (sPV > 1.6×10^{-4} s^{-1}) or both outside (sPV < 1.2×10^{-4} s^{-1}) the polar vortex at each layer. The time period selected for the comparisons with dynamical coincidence criteria was spring

(up to day 105), as defined in Section 4.1. This period includes all the springtime ACE measurements, and most of the days when the lower stratosphere might be inside the polar vortex (as indicated by the DMP calculations).

For ozone, altitude levels of 14, 18, 20, and 22 km were selected, to coincide with the location of the peak ozone concentrations in the lower stratosphere. Adams et al. [8] used similar altitude levels (in pressure coordinates), excluding 22 km, while Batchelor et al. [10] and Griffin et al. [12] used the same levels with the addition of 24, 26, 30, 36, and 46 km. Imposing dynamical coincidence criteria throughout the entire stratosphere drastically reduced comparison statistics due to the 500 km distance limit used in this study. With the additional coincidence criteria, and using surface-52 km ozone columns, the ACE minus Bruker FTIR and ACE minus PARIS-IR comparisons showed large improvements. For the Bruker FTIR, ACE-FTS relative differences improved from $-7.4 \pm 0.2\%$ to $-3.9 \pm 0.4\%$, while ACE-MAESTRO comparisons improved from $-12.0 \pm 0.3\%$ to $-7.3 \pm 0.6\%$. For the PARIS-IR, the changes were $-4.3 \pm 0.1\%$ to $-2.0 \pm 0.3\%$, and $-8.8 \pm 0.1\%$ to $-4.5 \pm 0.5\%$, respectively. Comparisons of ACE-FTS and ACE-MAESTRO measurements to ZSL-DOAS data did not improve with the additional coincidence criteria. OSIRIS comparisons to the ZSL-DOAS measurements showed modest improvements, from $6.7 \pm 0.2\%$ to $5.9 \pm 0.2\%$ for GBS and $2.3 \pm 0.2\%$ to $1.8 \pm 0.2\%$ for SAOZ. OSIRIS minus Bruker FTIR and PARIS-IR comparisons did not change significantly with the inclusion of the dynamical coincidence criteria.

The comparison results with the stricter coincidence criteria are similar to the findings of Adams et al. [8]. They also saw the largest improvement for ACE-FTS minus Bruker FTIR (-5.0% to -3.1%), with modest or no improvements for the other instrument pairs. The good agreement of the ACE-FTS to Bruker FTIR comparisons using the dynamical coincidence criteria ($-3.1 \pm 0.8\%$ [8], $-3.6 \pm 0.6\%$ [12], and $-3.9 \pm 0.4\%$ here) indicates that year-to-year variability in the location and strength of the polar vortex has a significant impact on the observed differences, and that the underlying agreement between ACE-FTS and Bruker FTIR is stable over time. There are no significant drifts in any of the relative difference time series filtered by the dynamical coincidence criteria.

While the location of the polar vortex influences the comparison results, removing these impacts is challenging for scattered light measurements such as OSIRIS and the ZSL-DOAS instruments. Precise line-of-sight calculations are not possible for these instruments due to the multiple paths taken by scattered sunlight before reaching the detectors. The dynamical coincidence criteria for OSIRIS are weakened due to the lack of line-of-sight information. In addition, ZSL-DOAS measurements are integrated over a much longer time than solar measurements. Springtime vertical columns are calculated using 2–4 h of measurements, corresponding to a 30–60° change in solar azimuth. Using the estimated line-of-sight for the mean measurement time only [8] may reduce the utility of the stricter coincidence criteria. Modest to no improvements in comparisons involving scattered light instruments are thus likely partially due to the cruder line-of-sight estimates used.

To investigate the impact of the polar vortex on 12–40 km NO_2 partial column comparisons, altitude levels of 24, 26, and 30 km were selected, since the peak NO_2 concentrations occur at higher altitudes than for ozone. The dynamical coincidence criteria, however, did not improve the comparison results, with the exception of the OSIRIS minus SAOZ relative differences ($-12.2 \pm 0.9\%$ to $-7.6 \pm 1.1\%$). The small impact of the dynamical coincidence criteria for NO_2 is likely due to the large uncertainties in the measurements and the diurnal scaling, as well as the variability of NO_2, such as the latitudinal gradient and the diurnal effect [4,8].

To assess the impact of the latitudinal gradient on NO_2 comparisons, a $\pm 1°$ latitude coincidence criterion was implemented using the 25 km tangent height of the OSIRIS measurements, the 30 km tangent height of the ACE-FTS measurements, and the corresponding 25 or 30 km point along the line-of-sight of the ground-based instruments. Since the latitudinal gradient is present during both spring and fall, measurements from both seasons (as defined in Section 4.1) were included. Ratios of modeled NO_2 partial columns at twilight for 79°N over 81°N (after Fig. 5 of Adams et al. [8]) are only as high as 2.5, compared to ratios of 7 for latitude differences typical for coincidences within a 500 km radius. The latitude filter improved the relative and absolute differences for six of the eight instrument pairs, although these improvements are only significant within standard error for the ACE-FTS minus GBS-UV comparison ($8.8 \pm 1.1\%$ to $2.2 \pm 2.6\%$). The remaining two comparisons (ACE-FTS to SAOZ and Bruker FTIR) resulted in inconclusive changes, where only the absolute or relative differences improved (none significant within standard error). The small changes using the $\pm 1°$ latitude coincidence criterion indicate that using the 500 km radius to select coincidences is adequate even when NO_2 has a strong latitudinal gradient. If a 1000 km radius is used, the $\pm 1°$ criterion significantly changes the results for six out of the eight instrument pairs, indicating that the larger radius would lead to systematic issues in the spring and fall NO_2 comparisons.

To estimate the impact of the diurnal effect, a new set of diurnal scaling factors was calculated for the ground-based instruments, using the SZA of the 30 km point along the line-of-sight instead of the time of the measurements. As expected, the scaled NO_2 partial columns increased, by 6–7% for the ZSL-DOAS instruments, and by 4% for the Bruker FTIR. The corresponding shift in the satellite minus ground-based differences results in better agreement for all the pairs where the satellite instrument overestimated the ground-based measurements, and the differences increased for pairs where the satellite instrument was already underestimating NO_2 compared to the ground-based data (see Fig. 8). These results indicate that the variability of NO_2 has a significant impact on comparisons at high latitudes.

7.2. Cloud-filtered ZSL-DOAS ozone dataset

Clouds are a large factor of uncertainty in ZSL-DOAS measurements, and they are not taken into account in the NDACC ozone AMF calculations. Hendrick et al. [30] estimated that this omission accounts for a 3.3% uncertainty in the ozone columns. Cloudy AMFs are systematically larger than AMFs in clear conditions, mainly due to multiple scattering in the cloud layer. Existing cloud-screening algorithms for (mainly off-axis) DOAS instruments [66–69] are based on the color index (CI) at various wavelength pairs. These algorithms, however, require small SZA measurements that are not available at high latitudes.

To assess the impact of clouds on ZSL-DOAS ozone measurements at Eureka, a cloud screening algorithm was developed for the UT-GBS and SAOZ datasets by Zhao et al. [70]. Cloudy spectra were filtered out using calibrated CI and thresholds based on radiative transfer model simulations prior to the VCD retrieval. This method, however, still requires SZA values smaller than 85° for the individual measurements, and so has limited impact on the spring and fall measurements when only the UT-GBS daily reference spectra might be filtered. To extend the range of the filter, the temporal smoothness of the CI and the O_4 dSCDs was also taken into account, since they should vary smoothly in the absence of rapidly varying clouds. The cloud-filtered ZSL-DOAS datasets were retrieved using the GBS retrieval code (Section 3.1), with the exception of the use of fixed reference column densities for SAOZ. Zhao et al. [70] found that for 2010–2017 data, there is a 1–5% difference between cloudy and clear ZSL-DOAS measurements. Cloudy measurements show a positive bias, consistent with the enhanced AMFs in cloudy conditions.

Using the same 2010–2017 subset of GBS and SAOZ measurements as in Zhao et al. [70], the effects of clouds on the satellite minus ZSL-DOAS intercomparisons can be examined. The ACE-FTS and ACE-MAESTRO comparisons show only insignificant changes. This is consistent with the limited applicability of the cloud-filter algorithm in the spring and fall. When compared to OSIRIS data, the mean relative differences increased for both GBS and SAOZ, although the change is only significant within standard error for the former. Since the GBS dataset has more summer data than SAOZ, the effect of the cloud filter is expected to be larger. OSIRIS minus GBS differences changed from $3.2 \pm 0.1\%$ to $4.4 \pm 0.2\%$, while correlation coefficients improved from 0.92 to 0.95. This is consistent with Zhao et al. [70], who found that the cloud filter changes the GBS minus Brewer differences from $0.05 \pm 0.25\%$ to $-1.84 \pm 0.71\%$.

These results suggest that clouds play an important role in satellite comparisons to ZSL-DOAS instruments at high latitudes. While the impact of clouds on spring and fall measurements is difficult to quantify, it is likely that ZSL-DOAS measurements have a positive bias during those periods as well. Since direct sun measurements (e.g. the Bruker FTIR and Brewer measurements) have a natural clear-sky bias, comparison results across multiple datasets have to be interpreted with care.

8. Conclusions

OSIRIS and ACE ozone and NO_2 measurements within 500 km of the PEARL Ridge Lab were compared to ground-based measurements. Ozone partial columns from 14 to 52 km were calculated from the satellite profiles, and these were extended to the surface using ozonesonde data. NO_2 partial columns were calculated from 12 to 40 km for ACE-FTS and 12 to 32 km for OSIRIS (scaled to 40 km using the NDACC NO_2 profile climatology). NO_2 partial columns were not extended to the surface since the ground-based instruments measure partial columns above 12 km. All NO_2 measurements were scaled to local noon using a photochemical model to account for the diurnal variation of NO_2. Drifts between the various datasets were calculated using robust linear regression of the daily mean relative differences.

Ozone partial columns from the three satellite instruments show reasonable agreement. OSIRIS and ACE-FTS agree to within 1.2%, while ACE-MAESTRO ozone shows a 6.7% and 5.9% low bias when compared to OSIRIS and ACE-FTS, respectively. Profile comparisons show that relative to the OSIRIS and ACE-FTS, ACE-MAESTRO underestimates the peak ozone concentrations within 500 km of PEARL. This bias was not apparent in previous ACE-MAESTRO data versions [8]. The lack of drift between any of the satellite datasets indicates that the ACE-MAESTRO bias is related to changes in the v3.13 processing, and not to changes in the dataset over time. While we cannot draw any conclusions regarding the global ACE-MAESTRO data, we advise caution when using the v3.13 ozone dataset in the Arctic for applications where accuracy is important. Satellite NO_2 partial columns were not compared due to the low number of coincidences betewen OSIRIS and ACE-FTS.

Satellite-plus-sonde ozone columns were compared to five ground-based datatsets. OSIRIS ozone columns agree with ground-based total columns with a maximum mean relative difference of 4.4%. The agreement is better than 7.5% for ACE-FTS ozone, while ACE-MAESTRO columns show a maximum relative difference of 12%, reflecting the low bias indicated by the satellite comparisons. The largest differences were observed for the ACE minus Bruker FTIR and PARIS-IR comparisons. Exluding those four values, all other instrument pairs agree to within 4.4%. Comparisons of satellite-plus-sonde ozone with a cloud-filtered ZSL-DOAS dataset indicate that the underlying agreements are likely different due to a positive bias in the ZSL-DOAS measurements in cloudy conditions.

Springtime ozone comparisons are affected by the polar vortex. This additional atmospheric variability is most significant for the ACE minus Bruker FTIR and PARIS-IR comparisons, since most of those coincidences (all for PARIS-IR) occur during the spring. Using stricter dynamical coincidence citeria in the spring, ACE minus Bruker FTIR and PARIS-IR comparisons improved by 2.3–4.7%. Results for the other instrument pairs showed modest to no improvements, likely due to the lack of precise line-of-sight information for the scattered-light measurements. This indicates that the polar vortex introduces significant uncertainty in the springtime comparisons, and accounting for these effects requires precise knowledge of the measurement light path.

Satellite NO_2 partial columns were compared to four ground-based datasets. OSIRIS partial columns agree with ground-based partial columns to within 19.9%, and the differences show significant seasonal variation, with the largest negative values in the summer. ACE-FTS partial columns show a maximum mean relative difference of 33.2%, that improves to better than 15.3% when excluding the Bruker FTIR comparison. Dynamical coincidence criteria did not improve the comparison results, likely due to the large uncertainties in the NO_2 measurements. Implementing a $\pm 1^\circ$ latitude coincidence criterion modestly improved spring and fall comparison results for most instrument pairs, suggesting that while the latitudinal gradient of NO_2 has a significant impact on validation exercises, a 500 km radius for coincidences is adequate for comparisons of NO_2 measurements.

None of the satellite-plus-sonde ozone columns or satellite NO_2 partial columns show a significant drift when compared to the ground-based datasets. While some of the calculated drifts are significant based on the linear regression only, most of these values were found to be related to the changing seasonal distribution of the coincidences, coupled with seasonal effects in the relative difference time series. None of the time series are long enough for confident detection of drifts of the magnitude given by the linear regression.

The results in this study are generally consistent with previous validation results (Fig. 9), with the exceptions as explained in Sections 5.1, 5.3, and 6.2. The lack of drift in any of the comparison time series indicates that OSIRIS, ACE-FTS and ACE-MAESTRO continue to provide reliable measurements of ozone and NO_2 in the Arctic.

Acknowledgements

The 2006–2017 UT-GBS, PEARL-GBS, SAOZ, Bruker FTIR and PARIS-IR measurements were made at PEARL by CANDAC. CANDAC has been supported by the Atlantic Innovation Fund/Nova Scotia Research Innovation Trust, Canada Foundation for Innovation, Canadian Foundation for Climate and Atmospheric Sciences (CFCAS), Canadian Space Agency (CSA), Environment and Climate Change Canada (ECCC), Government of Canada International Polar Year funding, Natural Sciences and Engineering Research Council (NSERC), Northern Scientific Training Program (NSTP), Ontario Innovation Trust, Polar Continental Shelf Program, and Ontario Research Fund. Brewer and ozonesonde measurements were made by ECCC. The spring 2004–2017 UT-GBS, PEARL-GBS, SAOZ, Bruker FTIR, PARIS-IR, Brewer, and ozonesonde measurements were also supported by the Canadian Arctic ACE/OSIRIS Validation Campaigns, which were funded by CSA, NSERC, NSTP, and ECCC. Spring 2007 GBS measurements were also supported by the Centre for Global Change Science. The 2001–2003 GBS measurements were supported by CFCAS and NSTP. K. Bognar was partially supported by the NSERC CREATE Training Program in Arctic Atmospheric Science, and the Arctic Validation And Training for Atmospheric Research in Space program, funded by CSA. SAOZ partic-

ipation in the campaigns was supported by the Centre National D'Etudes Spatiales. The authors wish to thank PEARL site manager Pierre Fogal, the CANDAC operators, and the staff at ECCC's Eureka Weather Station for their contributions to data acquisition, and logistical and on-site support. The authors also wish to thank Matthew Bassford, Elham Farahani, Annemarie Fraser, Cristen Adams, and Sophie Tran for their contribution to the GBS measurements, Rodica Lindenmaier, Rebecca Batchelor, Dan Weaver, Joseph Mendonca, and Sébastien Roche for their contribution to the Bruker FTIR measurements, and Deijan Fu, Ashley Harrett, Felicia Kolonjari, Rebecca Batchelor, Lin Dan, Ellen Eckert, Keeyoon Sung, and Dan Weaver for their contribution to the PARIS-IR measurements.

Odin is a Swedish-led satellite project funded jointly by Sweden (SNSB), the CSA, Finland (TEKES), France (CNES), and supported since 2007 by the third party mission programme of the European Space Agency. The Atmospheric Chemistry Experiment, also known as SCISAT, is a Canadian-led mission mainly supported by the CSA and NSERC. The authors wish to thank Peter Bernath for his leadership of the ACE mission. Work carried out at the Jet Propulsion Laboratory, California Institute of Technology was done under contract with the National Aeronautics and Space Administration.

Data availability: UT-GBS and PEARL-GBS ozone and NO_2-vis data, as well as the Bruker FTIR ozone profiles are available from the NDACC database hosted by NOAA at http://www.ndsc.ncep.noaa.gov/. UT-GBS and PEARL-GBS NO_2-UV data, as well as Bruker NO_2 profiles and PARIS-IR ozone profiles, are available upon request from the authors. The SAOZ ozone and NO_2 data can be found at http://saoz.obs.uvsq.fr/SAOZ_consol_v3.html. The NDACC AMF, profile, and averaging kernel look-up tables used in the ZSL-DOAS retrievals are provided by the NDACC UV–vis Working Group (http://ndacc-uvvis-wg.aeronomie.be/tools.php). The pressure and temperature profiles used in the ground-based FTIR retrievals are available from NCEP at http://ftp.cpc.ncep.noaa.gov/ndacc/ncep/. Ozonesonde and Brewer data are available from the World Ozone and Ultraviolet Radiation Data Centre (https://woudc.org). ACE-FTS and ACE-MAESTRO data are available after registration at https://databace.scisat.ca. OSIRIS data are available from the ESA third party mission database and on ftp://odin-osiris.usask.ca/. MERRA-2 data used for the DMP calculations are available at https://disc.sci.gsfc.nasa.gov/uui/datasets?keywords=MERRA-2.

Appendix A. Absolute differences as a function of solar zenith angle

Since seasonal plots of absolute differences might mask SZA-related patterns in the comparison results, Figs. A.10 and A.11 show absolute differences between satellite-plus-sonde ozone columns and the ground-based datasets as a function of SZA, while Fig. A.12 shows the same for satellite NO_2 partial columns. The SZA at the location of the ground-based instrument (and at the time of the ground-based measurement) was used in the figures. The mean absolute and relative differences, as well as the *RMSD* values, are the same as shown in Figs. 3, 5, and 8.

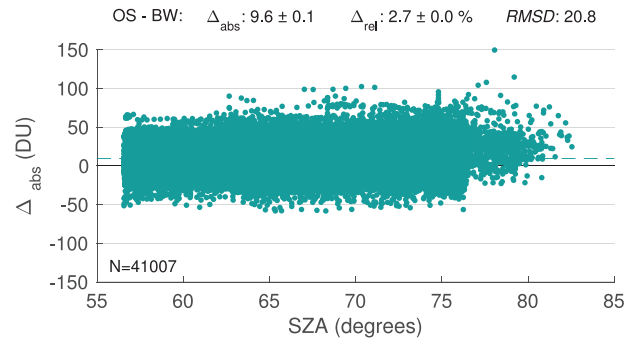

Fig. A.10. As for Fig. 3b, seasonal absolute differences between OSIRIS-plus-sonde surface-52 km ozone columns and Brewer total columns, as a function of the SZA corresponding to each Brewer measurement.

OSIRIS-plus-sonde ozone columns show no SZA-dependent differences when compared to Brewer ozone data (Fig. A.10). Brewer measurements only extend above SZA=76° for a few years, so the distribution of the differences for those angles is not well sampled. Removing Brewer measurements with SZA > 76° does not affect the comparison results. The OSIRIS minus Bruker FTIR differences (Fig. A.11b) are similarly independent of SZA when the sun is high, while for SZA > 80° the differences increase due to atmospheric variability in the spring. The OSIRIS minus PARIS-IR comparisons show the same pattern, with slightly more scatter in the absolute differences.

The satellite-plus-sonde minus ZSL-DOAS ozone differences (Fig. A.11a, c, and e) show no obvious SZA dependence. Most datapoints cluster around 88.5°, the mean SZA of the ideal twilight ZSL-DOAS measurements. The large scatter in these comparisons reflects the large differences observed in the spring. The ACE instruments show large scatter when compared to the Bruker FTIR and PARIS-IR ozone columns (Fig. A.11d, e), with the largest negative outliers concentrated around SZA ≳ 84°.

Most of the satellite NO_2 datasets show no obvious dependence on SZA when compared to the ground-based datasets (Fig. 8). The exception is the OSIRIS minus GBS-vis and GBS-UV, where the differences show large negative values for low SZA, reflecting the seasonal pattern described in Section 6.1.

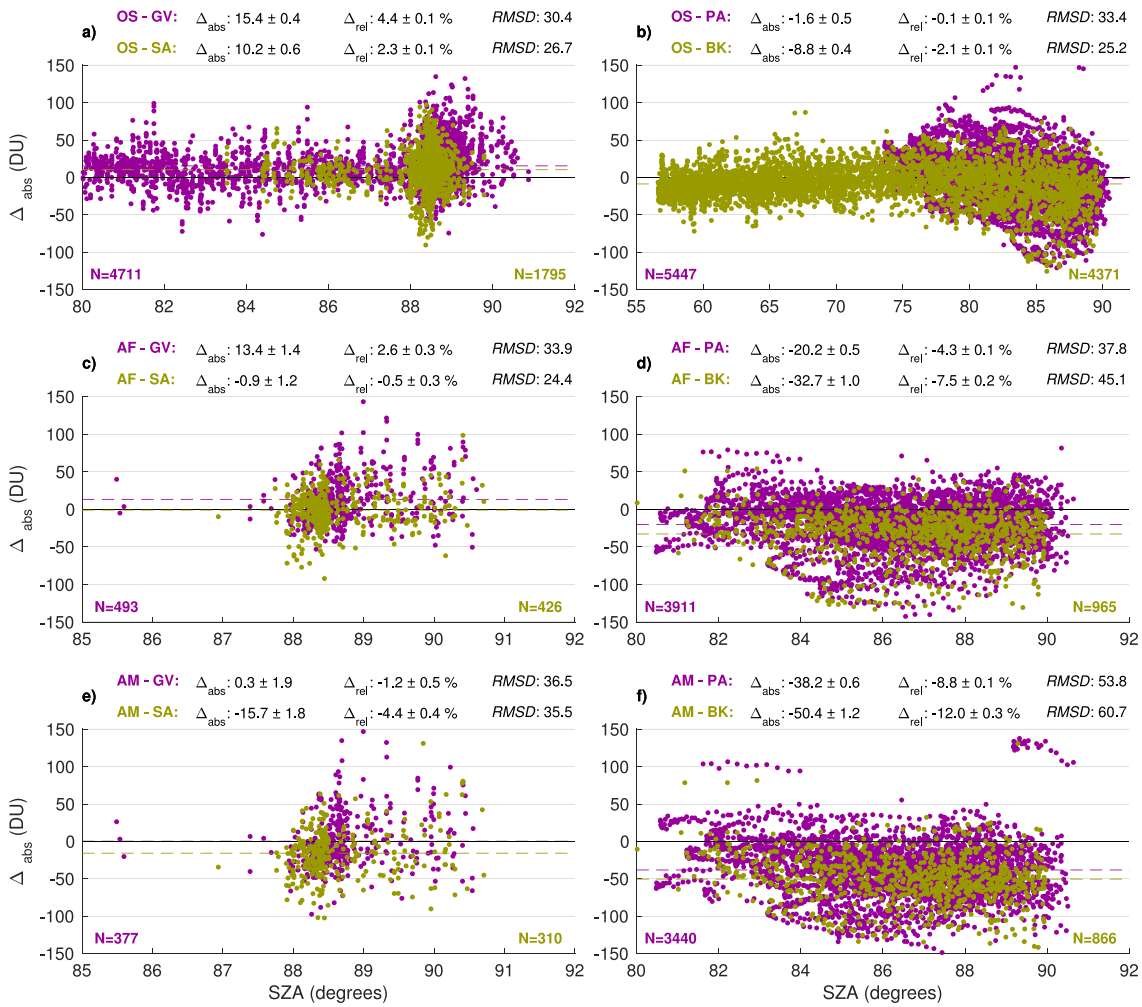

Fig. A.11. As for Fig. 5, with the SZA of the ground-based measurements on the *x*-axis instead of day of the year. Note that the SZA limits are different for each subplot.

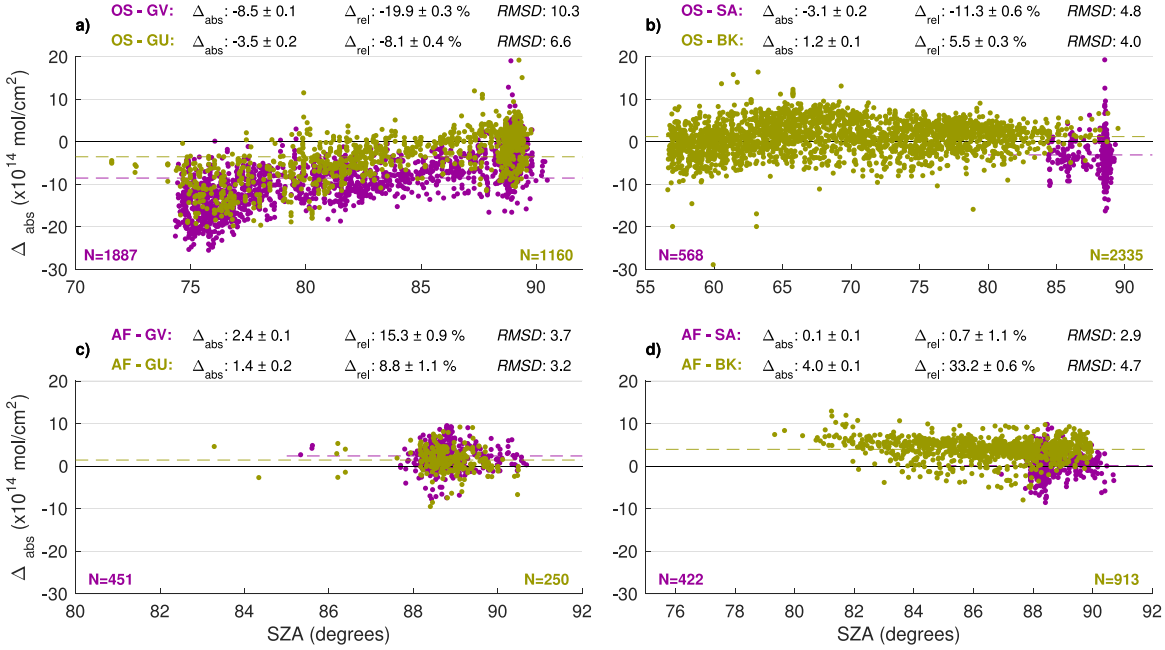

Fig. A.12. As for Fig. 8, 12–40 km NO_2 satellite partial columns against ground-based partial columns, with the SZA of the ground-based measurements on the *x*-axis instead of day of the year. Note that the SZA limits are different for each subplot.

Appendix B. Ground-based instrument intercomparisons

Understanding the relationship of the ground-based datasets with each other is important for interpreting the results of the satellite vs. ground-based instrument comparisons. Results for the ground-based ozone and NO_2 comparisons are summarised in the following sections. The coincidence criteria and comparison metrics were the same as detailed in Section 4.

B.1. Ozone

Ozone total columns from the GBS, SAOZ, Bruker FTIR, PARIS-IR and Brewer instruments show good correlation (Fig. B.13), with correlation coefficients of 0.83 or greater. The lowest value is found for the PARIS-IR minus Brewer comparison, since the two instruments have limited temporal overlap. The seasonal differences between the ground-based instruments are shown in Fig. B.14. The ZSL-DOAS instruments correlate very well (R=0.97), but the GBS ozone is systematically smaller than SAOZ, with a mean relative difference of -4.3%. The largest absolute differences occur in the spring. Comparing GBS to $SAOZ_{allyear}$ ozone (Section 3.1), the mean relative difference is reduced to -3.2%, due to better agreement in the summer. The mean difference of −4.3% is comparable to Fraser et al. [9] and Adams et al. [8], who found values of −6.9% to −3.7% (2005–2006), and −3.2% (2005–2011), respectively.

Both GBS and SAOZ compare well to Brewer ozone columns, with mean relative differences of −0.9% and −0.2%, respectively, indicating that outside early spring, ZSL-DOAS retrievals are consistent between GBS and SAOZ. These values also agree well with the results of Adams et al. [8], who found mean relative differences of −1.4% and 0.4% for GBS and SAOZ total columns, respectively.

The Bruker FTIR and PARIS-IR measure more ozone than the GBS, SAOZ, and Brewer instruments. The mean relative differences for the Bruker FTIR are 8.0%, 6.0%, and 4.5%, respectively, while the PARIS-IR differences are 9.3%, 3.7%, and 1.5%, respectively. The absolute differences are the largest in the early spring, and the overestimation by Bruker FTIR is present consistently during the entire year. These results for the Bruker FTIR are comparable to the values of 6.9%, 9.2%, and 2.6% from Adams et al. [8]. The Bruker FTIR minus Brewer relative differences agree well with Schneider et al. [71], who also found a 4.5% difference between the datasets at a subtropical site. Schneider et al. [71] attributed the differences to discrepancies in the UV and infrared spectroscopic parameters. The Bruker FTIR total columns show good agreement with PARIS-IR data, with a mean difference of −1.8%. This is consistent with Batchelor et al. [10] who found −1.2%, and Griffin et al. [12] who found −0.3%.

B.2. NO_2

The GBS-vis, GBS-UV, SAOZ, and Bruker FTIR NO_2 data show correlation coefficients of 0.92 or greater, as shown in Fig. B.15. The seasonal differences between the partial columns (12–60 km for the DOAS instruments and 12–40 km for the Bruker FTIR) are

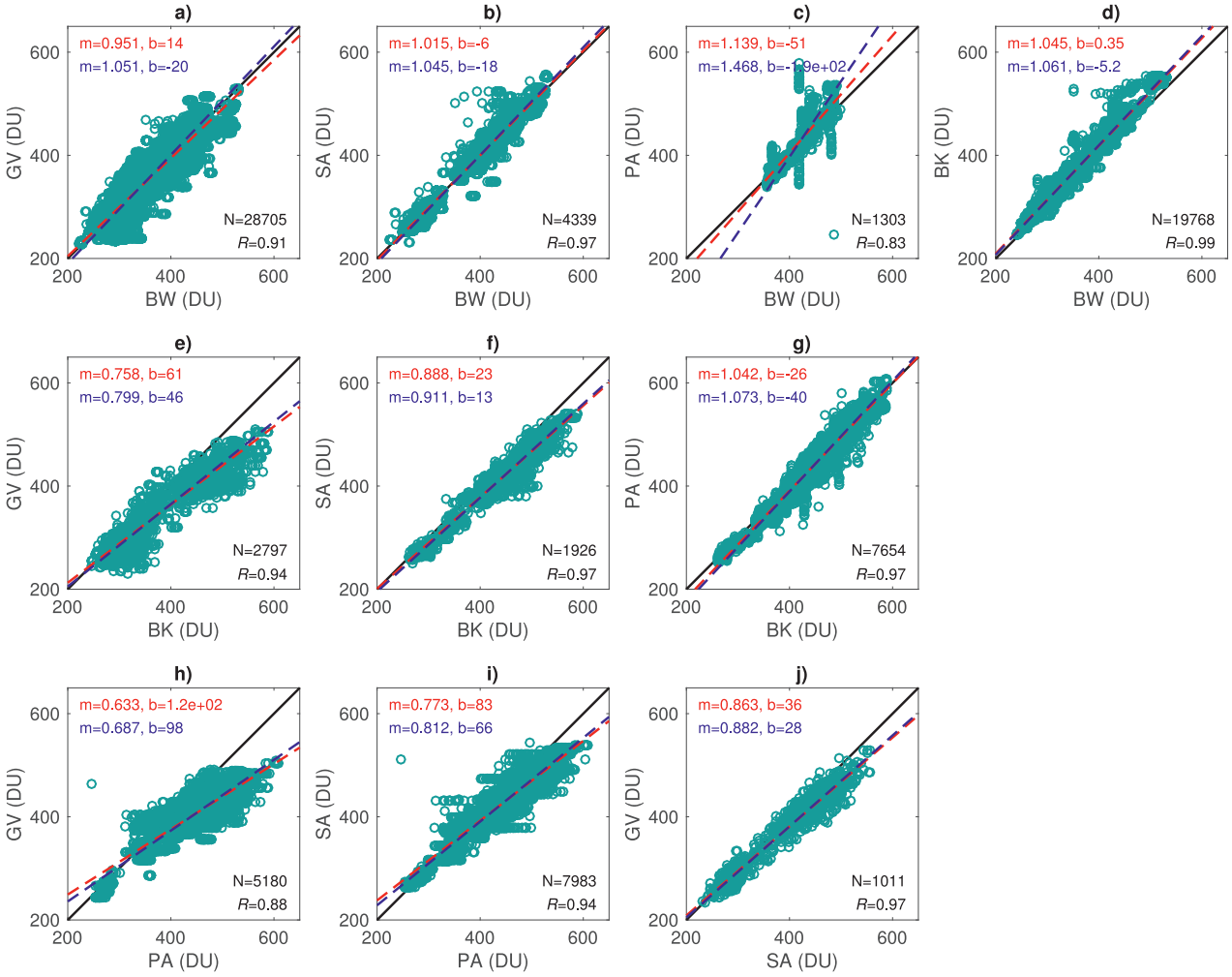

Fig. B.13. As for Fig. 4, ozone total columns from the ground-based instruments.

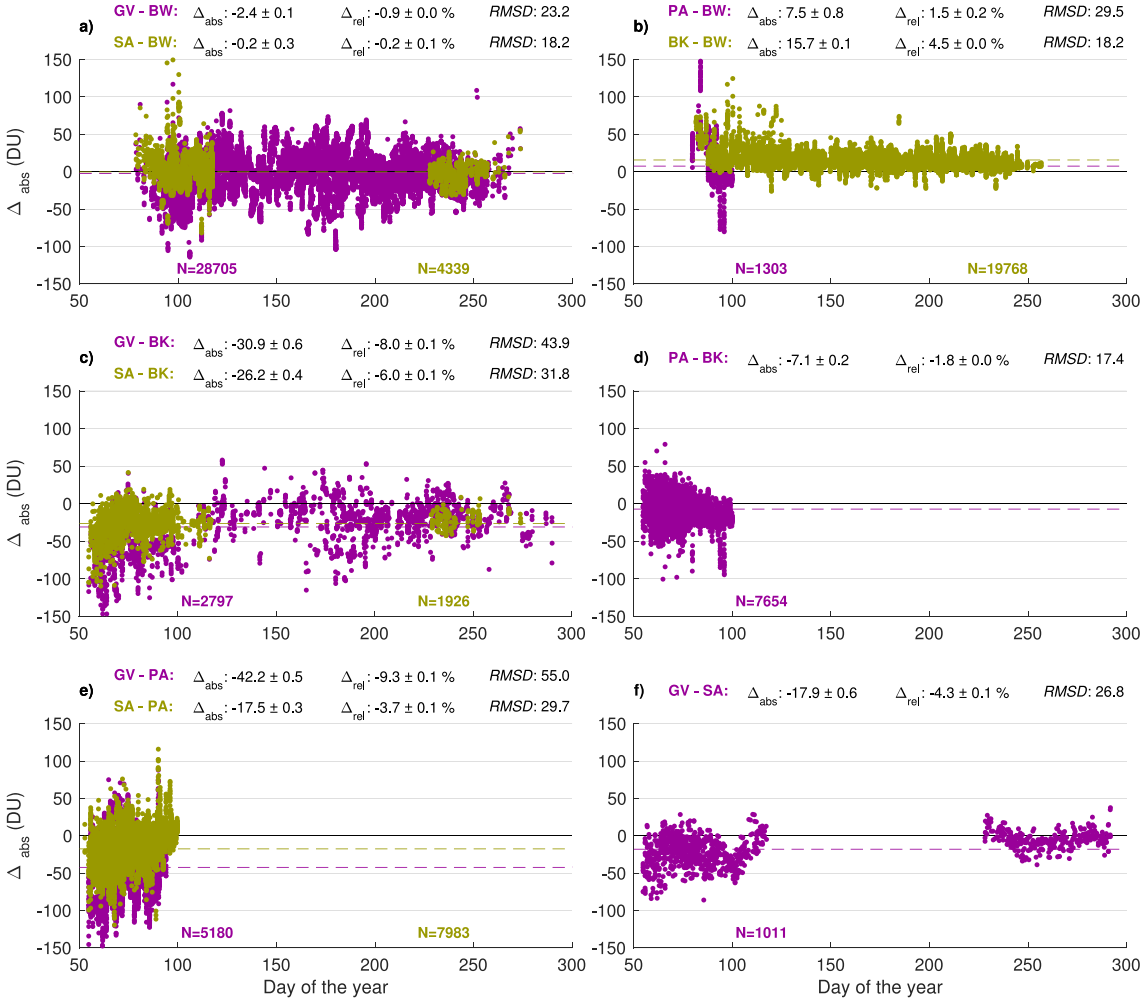

Fig. B.14. As for Fig. 5, ozone total columns from the ground-based instruments.

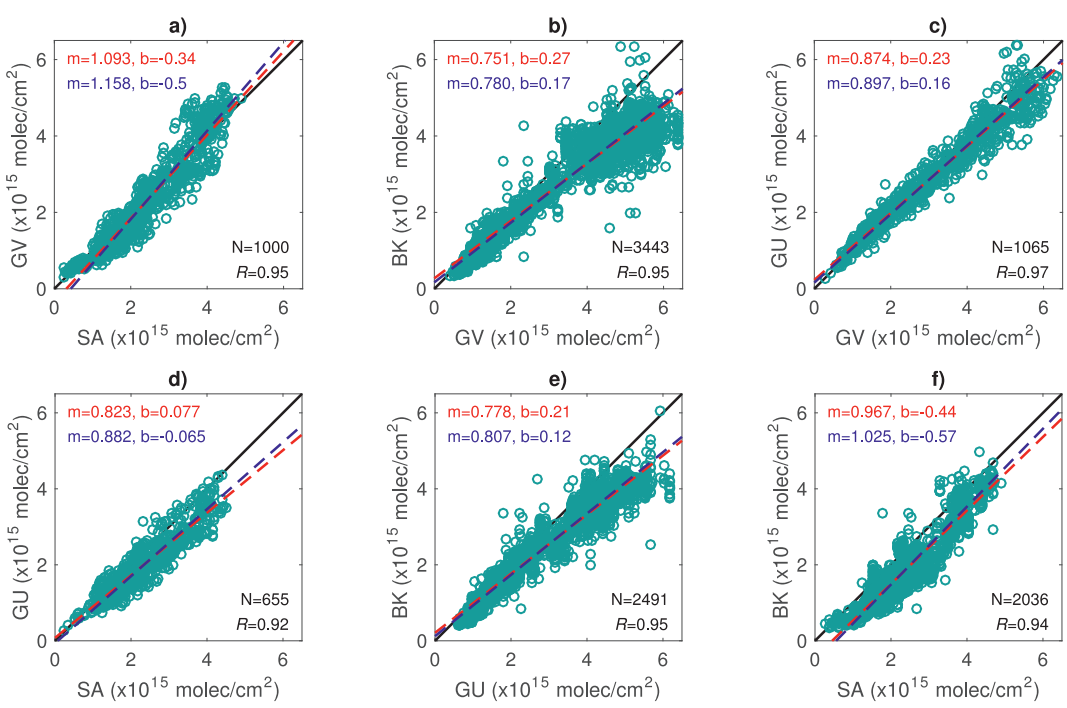

Fig. B.15. As for Fig. 4, 12–60 km NO_2 partial columns (12–40 km for the Bruker FTIR) from the ground-based instruments.

Please cite this article as: K. Bognar, X. Zhao and K. Strong et al., Updated validation of ACE and OSIRIS ozone and NO_2 measurements in the Arctic using ground-based instruments at Eureka, Canada, Journal of Quantitative Spectroscopy & Radiative Transfer, https://doi.org/10.1016/j.jqsrt.2019.07.014

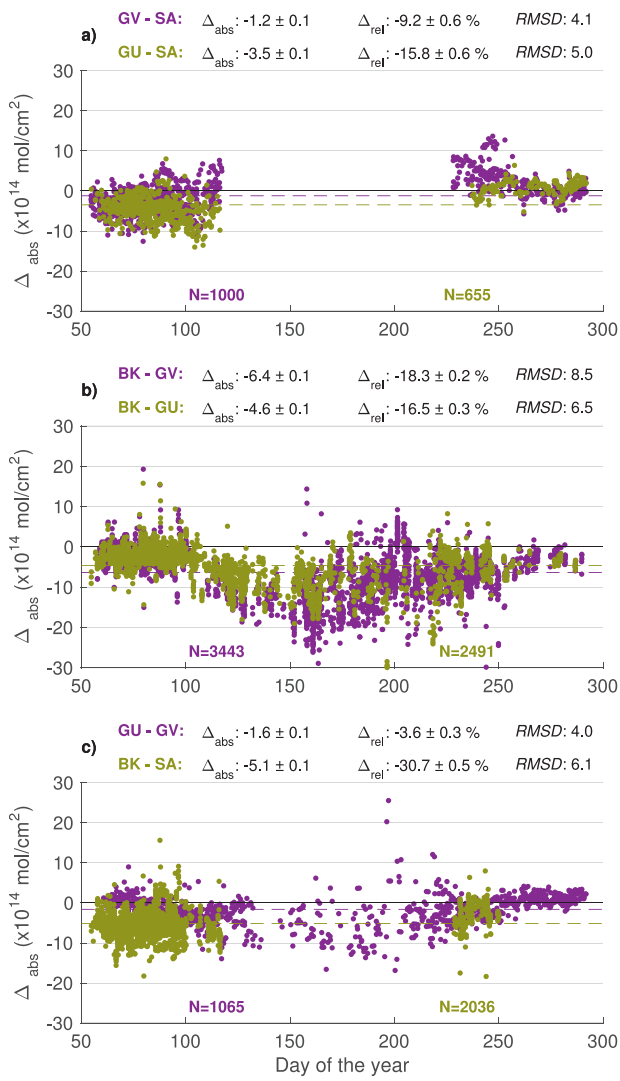

Fig. B.16. As for Fig. 5, 12–60 km NO_2 partial columns (12–40 km for the Bruker FTIR) from the ground-based instruments.

shown in Fig. B.16. The GBS-UV NO_2 is on average 3.6% lower than the GBS-vis product. This agreement is reasonable, given the shorter pathlengths corresponding to UV measurements. The GBS-vis and GBS-UV measurements are both smaller than SAOZ, with mean relative differences of −9.2% and −15.8%, respectively. The absolute differences are largest in the spring for both instrument pairs. The differences between SAOZ and the GBS datasets improve when the $SAOZ_{allyear}$ data are considered. GBS-vis minus $SAOZ_{allyear}$ relative differences change sign in the summer, leading to a large improvement in the mean, to −1.1%. GBS-UV partial columns are consistently smaller than $SAOZ_{allyear}$, with a mean relative difference of −11.6%. The results for the GBS-vis minus SAOZ comparisons are comparable to Fraser et al. [9], who found values of −2.2% to −12.2% for 2005–2006, using 22–40 km partial columns and identical software for the GBS-vis and SAOZ retrievals. Adams et al. [8] found mean differences of 3.8% and −6.4% for GBS-vis and GBS-UV compared to SAOZ using 17–40 km partial columns. The large difference compared to the results of this study is likely due to the changes in the SAOZ NO_2 dataset (see Section 3.1).

The Bruker FTIR NO_2 partial columns are smaller than all the ZSL-DOAS datasets, with mean relative differences of −18.3%, −16.5%, and −30.7% with respect to GBS-vis, GBS-UV, and SAOZ. The absolute differences peak in the summer, when NO_2 concentrations are at a maximum. The relative differences don't show any obvious seasonal variation. The smaller Bruker FTIR partial columns are consistent with the results of Adams et al. [8], who found mean relative differences of −16.3%, −19.2%, and −12.0% for GBS-vis, GBS-UV, and SAOZ, using a different version of the Bruker FTIR NO_2 retrievals. Part of the reason for the smaller Bruker FTIR measurements is the choice of partial column altitude range. When the upper limit of the partial columns is extended to 60 km to match the ZSL-DOAS retrievals, the mean relative differences improve to −10.6%, −8.5%, and −20.7% for GBS-vis, GBS-UV and SAOZ. The 12–40 km Bruker FTIR partial columns were used for the ground-based comparisons in order to keep the results consistent with the satellite comparisons.

Supplementary material

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.jqsrt.2019.07.014.

References

[1] Brohede SM, Haley CS, Mclinden CA, Sioris CE, Murtagh DP, Petelina SV, et al. Validation of Odin/OSIRIS stratospheric NO_2 profiles. J Geophys Res Atmos 2007;112(D7). doi:10.1029/2006JD007586.
[2] Degenstein DA, Bourassa AE, Roth CZ, Llewellyn EJ. Limb scatter ozone retrieval from 10 to 60 km using a multiplicative algebraic reconstruction technique. Atmos Chem Phys 2009;9(17):6521–9. doi:10.5194/acp-9-6521-2009.
[3] Dupuy E, Walker KA, Kar J, Boone CD, McElroy CT, Bernath PF, et al. Validation of ozone measurements from the atmospheric chemistry experiment (ACE). Atmos Chem Phys 2009;9(2):287–343. doi:10.5194/acp-9-287-2009.
[4] Kerzenmacher T, Wolff MA, Strong K, Dupuy E, Walker KA, Amekudzi LK, et al. Validation of NO_2 and no from the atmospheric chemistry experiment (ACE). Atmos Chem Phys 2008;8(19):5801–41. doi:10.5194/acp-8-5801-2008.
[5] Sioris C, Rieger L, Lloyd N, Bourassa A, Roth C, Degenstein D, et al. Improved OSIRIS NO_2 retrieval algorithm: description and validation. Atmos Meas Tech 2017;10(3):1155–68. doi:10.5194/amt-10-1155-2017.
[6] Fogal PF, LeBlanc LM, Drummond JR. The polar environment atmospheric research laboratory (PEARL): sounding the atmosphere at 80° North. ARCTIC 2013;66(3):377–86. doi:10.2307/23594645.
[7] Kerzenmacher TE, Walker KA, Strong K, Berman R, Bernath PF, Boone CD, et al. Measurements of O_3, NO_2 and temperature during the 2004 Canadian arctic ace validation campaign. Geophys Res Lett 2005;32(16):L16S07. doi:10.1029/2005GL023032.
[8] Adams C, Strong K, Batchelor RL, Bernath PF, Brohede S, Boone C, et al. Validation of ace and OSIRIS ozone and NO_2 measurements using ground based instruments at 80° N. Atmos Meas Tech 2012;5(5):927–53. doi:10.5194/amt-5-927-2012.
[9] Fraser A, Goutail F, Strong K, Bernath PF, Boone C, Daffer WH, et al. Intercomparison of uv-visible measurements of ozone and NO_2 during the Canadian arctic ace validation campaigns: 2004–2006. Atmos Chem Phys 2008;8(6):1763–88. doi:10.5194/acp-8-1763-2008.
[10] Batchelor RL, Kolonjari F, Lindenmaier R, Mittermeier RL, Daffer W, Fast H, et al. Four fourier transform spectrometers and the arctic polar vortex: instrument intercomparison and ACE-FTS validation at eureka during the IPY springs of 2007 and 2008. Atmos Meas Tech 2010;3(1):51–66. doi:10.5194/amt-3-51-2010.
[11] Fu D, Walker KA, Mittermeier RL, Strong K, Sung K, Fast H, et al. Simultaneous trace gas measurements using two fourier transform spectrometers at Eureka, Canada during spring 2006, and comparisons with the ACE-FTS. Atmos Chem Phys 2011;11(11):5383–405. doi:10.5194/acp-11-5383-2011.
[12] Griffin D, Walker KA, Conway S, Kolonjari F, Strong K, Batchelor R, et al. Multi-year comparisons of ground-based and space-borne fourier transform spectrometers in the high arctic between 2006 and 2013. Atmos Meas Tech 2017;10(9):3273–94. doi:10.5194/amt-10-3273-2017.
[13] Lindenmaier R, Strong K, Batchelor RL, Bernath PF, Chabrillat S, Chipperfield MP, et al. A study of the arctic NO_y budget above Eureka, Canada. J Geophys Res Atmos 2011;116(D23302). doi:10.1029/2011JD016207.
[14] Sung K, Skelton R, Walker KA, Boone CD, Fu D, Bernath PF. N_2O and O_3 arctic column amounts from paris-ir observations: retrievals, characterization and error analysis. J Quant Spectrosc Radiat Transf 2007;107(3):385–406. doi:10.1016/j.jqsrt.2007.03.002.
[15] Fraser A, Adams C, Drummond JR, Goutail F, Manney G, Strong K. The polar environment atmospheric research laboratory uv–visible ground-based spectrometer: first measurements of o_3, no_2, bro, and oclo columns. J Quant Spectrosc Radiat Transf 2009;110(12):986–1004. doi:10.1016/j.jqsrt.2009.02.034.
[16] Pommereau JP, Goutail F. O_3 and NO_2 ground-based measurements by visible spectrometry during arctic winter and spring 1988. Geophys Res Lett 1988;15(8):891–4. doi:10.1029/GL015i008p00891.
[17] Batchelor RL, Strong K, Lindenmaier R, Mittermeier RL, Fast H, Drummond JR, et al. A new bruker IFS 125HR FTIR spectrometer for the polar environment

atmospheric research laboratory at Eureka, Nunavut, Canada: measurements and comparison with the existing bomem DA8 spectrometer. J Atmos Ocean Tech 2009;26(7):1328–40. doi:10.1175/2009JTECHA1215.1.
[18] Fu D, Walker KA, Sung K, Boone CD, Soucy M-A, Bernath PF. The portable atmospheric research interferometric spectrometer for the infrared, PARIS-IR. J Quant Spectrosc Radiat Transf 2007;103(2):362–70. doi:10.1016/j.jqsrt.2006.05.006.
[19] Kerr JB. New methodology for deriving total ozone and other atmospheric variables from brewer spectrophotometer direct sun spectra. J Geophys Res Atmos 2002;107(D23). doi:10.1029/2001JD001227. ACH 22–1–ACH 22–17.
[20] Tarasick DW, Davies J, Smit HGJ, Oltmans SJ. A re-evaluated canadian ozonesonde record: measurements of the vertical distribution of ozone over Canada from 1966 to 2013. Atmos Meas Tech 2016;9(1):195–214. doi:10.5194/amt-9-195-2016.
[21] Murtagh D, Frisk U, Merino F, Ridal M, Jonsson A, Stegman J, et al. An overview of the odin atmospheric mission. Can J Phys 2002;80(4):309–19. doi:10.1139/p01-157.
[22] Llewellyn EJ, Lloyd ND, Degenstein DA, Gattinger RL, Petelina SV, Bourassa AE, et al. The OSIRIS instrument on the odin spacecraft. Can J Phys 2004;82(6):411–22. doi:10.1139/p04-005.
[23] Bourassa AE, Roth CZ, Zawada DJ, Rieger LA, McLinden CA, Degenstein DA. Drift-corrected Odin-OSIRIS ozone product: algorithm and updated stratospheric ozone trends. Atmos Meas Tech 2018;11(1):489–98. doi:10.5194/amt-11-489-2018.
[24] Zawada DJ, Dueck SR, Rieger LA, Bourassa AE, Lloyd ND, Degenstein DA. High-resolution and monte carlo additions to the sasktran radiative transfer model. Atmos Meas Tech 2015;8(6):2609–23. doi:10.5194/amt-8-2609-2015.
[25] Bernath PF, McElroy CT, Abrams MC, Boone CD, Butler M, Camy-Peyret C, et al. Atmospheric chemistry experiment (ACE): mission overview. Geophys Res Lett 2005;32(15):L15S01. doi:10.1029/2005GL022386.
[26] Boone CD, Nassar R, Walker KA, Rochon Y, McLeod SD, Rinsland CP, et al. Retrievals for the atmospheric chemistry experiment fourier-transform spectrometer. Appl Opt 2005;44(33):7218–31. doi:10.1364/AO.44.007218.
[27] Boone CD, Walker KA, Bernath PF. Version 3 retrievals for the atmospheric chemistry experiment fourier transform spectrometer (ACE-FTS). In: Bernath PF, editor. The atmospheric chemistry experiment ACE at 10: a solar occultation anthology. A. Deepak Publishing; 2013. p. 103–27.
[28] McElroy CT, Nowlan CR, Drummond JR, Bernath PF, Barton DV, Dufour DG, et al. The ACE-MAESTRO instrument on SCISAT: description, performance, and preliminary results. Appl Opt 2007;46(20):4341–56. doi:10.1364/AO.46.004341.
[29] Platt U, Stutz J. Differential optical absorption spectroscopy. Springer; 2008.
[30] Hendrick F, Pommereau J-P, Goutail F, Evans RD, Ionov D, Pazmino A, et al. NDACC/SAOZ UV-visible total ozone measurements: improved retrieval and comparison with correlative ground-based and satellite observations. Atmos Chem Phys 2011;11(12):5975–95. doi:10.5194/acp-11-5975-2011.
[31] Pougatchev NS, Connor BJ, Rinsland CP. Infrared measurements of the ozone vertical distribution above KITT peak. J Geophys Res Atmos 1996;100(D8):16689–97. doi:10.1029/95JD01296.
[32] Rodgers CD. Inverse methods for atmospheric sounding: theory and practice, 2. World Scientific; 2000. ISBN 981-02-2740-X.
[33] Eyring V, Waugh DW, Bodeker GE, Cordero E, Akiyoshi H, Austin J, et al. Multimodel projections of stratospheric ozone in the 21st century. J Geophys Res Atmos 2007;112(D16). doi:10.1029/2006JD008332.
[34] Rothman LS, Gordon IE, Barbe A, Benner DC, Bernath PF, Birk M, et al. The HITRAN 2008 molecular spectroscopic database. J Quant Spectrosc Radiat Transf 2009;110(9–10):533–72. doi:10.1016/j.jqsrt.2009.02.013.
[35] von Clarmann T. Smoothing error pitfalls. Atmos Meas Tech 2014;7(9):3023–34. doi:10.5194/amt-7-3023-2014.
[36] York D, Evensen NM, Martınez ML, De Basabe Delgado J. Unified equations for the slope, intercept, and standard errors of the best straight line. Am J Phys 2004;72(3):367–75. doi:10.1119/1.1632486.
[37] Gruber A, Su C-H, Zwieback S, Crow W, Dorigo W, Wagner W. Recent advances in (soil moisture) triple collocation analysis. Int J Appl Earth Obs Geoinform 2016;45:200–11. doi:10.1016/j.jag.2015.09.002.
[38] Kinzel J, Fennig K, Schröder M, Andersson A, Bumke K, Hollmann R. Decomposition of random errors inherent to HOAPS-3.2 near-surface humidity estimates using multiple triple collocation analysis. J Atmos Ocean Tech 2016;33(7):1455–71. doi:10.1175/JTECH-D-15-0122.1.
[39] Loew A, Schlenz F. A dynamic approach for evaluating coarse scale satellite soil moisture products. Hydrol Earth Syst Sci 2011;15(1):75–90. doi:10.5194/hess-15-75-2011.
[40] Loew A, Bell W, Brocca L, Bulgin CE, Burdanowitz J, Calbet X, et al. Validation practices for satellite-based earth observation data across communities. Rev of Geophys 2017;55(3):779–817. doi:10.1002/2017RG000562.
[41] McColl KA, Vogelzang J, Konings AG, Entekhabi D, Piles M, Stoffelen A. Extended triple collocation: estimating errors and correlation coefficients with respect to an unknown target. Geophys Res Lett 2014;41(17):6229–36. doi:10.1002/2014GL061322.
[42] Stoffelen A. Toward the true near-surface wind speed: error modeling and calibration using triple collocation. J Geophys Res Ocean 1998;103(C4):7755–66. doi:10.1029/97JC03180.
[43] Su C-H, Ryu D, Crow WT, Western AW. Beyond triple collocation: applications to soil moisture monitoring. J Geophys Res Atmos 2014;119(11):6419–39. doi:10.1002/2013JD021043.
[44] Verhoelst T, Granville J, Hendrick F, Köhler U, Lerot C, Pommereau J-P, et al. Metrology of ground-based satellite validation: co-location mismatch and smoothing issues of total ozone comparisons. Atmos Meas Tech 2015;8(12):5039–62. doi:10.5194/amt-8-5039-2015.
[45] Brohede S, McLinden CA, Berthet G, Haley CS, Murtagh D, Sioris CE. A stratospheric NO_2 climatology from Odin/OSIRIS limb-scatter measurements. Can J Phys 2007;85(11):1253–74. doi:10.1139/p07-141.
[46] McLinden CA, Olsen SC, Hannegan B, Wild O, Prather MJ, Sundet J. Stratospheric ozone in 3-D models: a simple chemistry and the cross-tropopause flux. J Geophys Res Atmos 2000;105(D11):14653–65. doi:10.1029/2000JD900124.
[47] Fish DJ, Jones RL, Strong EK. Midlatitude observations of the diurnal variation of stratospheric bro. J Geophys Res Atmos 1995;100(D9):18863–71. doi:10.1029/95JD01944.
[48] Hendrick F, Van Roozendael M, Kylling A, Petritoli A, Rozanov A, Sanghavi S, et al. Intercomparison exercise between different radiative transfer models used for the interpretation of ground-based zenith-sky and multi-axis DOAS observations. Atmos Chem Phys 2006;6(1):93–108. doi:10.5194/acp-6-93-2006.
[49] McLinden CA, Haley CS, Sioris CE. Diurnal effects in limb scatter observations. J Geophys Res Atmos 2006;111(D14302). doi:10.1029/2005JD006628.
[50] Newchurch MJ, Allen M, Gunson MR, Salawitch RJ, Collins GB, Huston KH, et al. Stratospheric no and NO_2 abundances from atmos solar-occultation measurements. Geophys Res Lett 1996;23(17):2373–6. doi:10.1029/96GL01196.
[51] Hubert D, Lambert J-C, Verhoelst T, Granville J, Keppens A, Baray J-L, et al. Ground-based assessment of the bias and long-term stability of 14 limb and occultation ozone profile data records. Atmos Meas Tech 2016;9(6):2497–534. doi:10.5194/amt-9-2497-2016.
[52] Adams C, Bourassa AE, Sofieva V, Froidevaux L, McLinden CA, Hubert D, et al. Assessment of Odin-OSIRIS ozone measurements from 2001 to the present using mls, gomos, and ozonesondes. Atmos Meas Tech 2014;7(1):49–64. doi:10.5194/amt-7-49-2014.
[53] Holland PW, Welsch RE. Robust regression using iteratively reweighted least-squares. Commun Stat Theory Methods 1977;6(9):813–27. doi:10.1080/03610927708827533.
[54] Efron B, Tibshirani R. Bootstrap methods for standard errors, confidence intervals, and other measures of statistical accuracy. Stat Sci 1986;1(1):54–75. doi:10.2307/2245500.
[55] Weatherhead EC, Reinsel GC, Tiao GC, Meng X-L, Choi D, Cheang W-K, et al. Factors affecting the detection of trends: statistical considerations and applications to environmental data. J Geophys Res Atmos 1998;103(D14):17149–61. doi:10.1029/98JD00995.
[56] Gardiner T, Forbes A, de Mazière M, Vigouroux C, Mahieu E, Demoulin P, et al. Trend analysis of greenhouse gases over europe measured by a network of ground-based remote FTIR instruments. Atmos Chem Phys 2008;8(22):6719–27. doi:10.5194/acp-8-6719-2008.
[57] Rodgers CD, Connor BJ. Intercomparison of remote sounding instruments. J Geophys Res Atmos 2003;108(D3). doi:10.1029/2002JD002299.
[58] Eskes HJ, Boersma KF. Averaging kernels for DOAS total-column satellite retrievals. Atmos Chem Phys 2003;3(5):1285–91. doi:10.5194/acp-3-1285-2003.
[59] Manney GL, Daffer WH, Zawodny JM, Bernath PF, Hoppel KW, Walker KA, et al. Solar occultation satellite data and derived meteorological products: sampling issues and comparisons with aura microwave limb sounder. J Geophys Res Atmos 2007;112(D24). doi:10.1029/2007JD008709.
[60] Gelaro R, McCarty W, Suárez MJ, Todling R, Molod A, Takacs L, et al. The modern-era retrospective analysis for research and applications, version 2 (MERRA-2). J Clim 2017;30(14):5419–54. doi:10.1175/JCLI-D-16-0758.1.
[61] GMAO. MERRA-2 inst3_3d_asm_Nv: 3d,3-hourly,instantaneous,model-level, assimilation,assimilated meteorological fields V5.12.4. Goddard Earth Sciences Data and Information Services Center (GES DISC); Greenbelt, MD, USA; 2015. doi:10.5067/WWQSXQ8IVFW8. Accessed: 2018/11/01.
[62] Dunkerton TJ, Delisi DP. Evolution of potential vorticity in the winter stratosphere of january-february 1979. J Geophys Res Atmos 1986;91(D1):1199–208. doi:10.1029/JD091iD01p01199.
[63] Manney GL, Zurek RW, O'Neill A, Swinbank R. On the motion of air through the stratospheric polar vortex. J Atmos Sci 1994;51(20):2973–94. doi:10.1175/1520-0469(1994)051<2973:OTMOAT>2.0.CO;2.
[64] Manney GL, Hegglin MI, Daffer WH, Santee ML, Ray EA, Pawson S, et al. Jet characterization in the upper troposphere/lower stratosphere (UTLS): applications to climatology and transport studies. Atmos Chem Phys 2011;11(12):6115–37. doi:10.5194/acp-11-6115-2011.
[65] Manney GL, Hegglin MI, Lawrence ZD, Wargan K, Millán LF, Schwartz MJ, et al. Reanalysis comparisons of upper tropospheric-lower stratospheric jets and multiple tropopauses. Atmos Chem Phys 2017;17(18):11541–66. doi:10.5194/acp-17-11541-2017.
[66] Gielen C, Van Roozendael M, Hendrick F, Pinardi G, Vlemmix T, De Bock V, et al. A simple and versatile cloud-screening method for MAX-DOAS retrievals. Atmos Meas Tech 2014;7(10):3509–27. doi:10.5194/amt-7-3509-2014.
[67] Wagner T, Apituley A, Beirle S, Dörner S, Friess U, Remmers J, et al. Cloud detection and classification based on MAX-DOAS observations. Atmos Meas Tech 2014;7(5):1289–320. doi:10.5194/amt-7-1289-2014.
[68] Wagner T, Beirle S, Remmers J, Shaiganfar R, Wang Y. Absolute calibration of the colour index and O_4 absorption derived from multi axis (MAX-) DOAS measurements and their application to a standardised cloud classification algorithm. Atmos Meas Tech 2016;9(9):4803–23. doi:10.5194/amt-9-4803-2016.
[69] Wang Y, Penning de Vries M, Xie P, Beirle S, Dörner S, Remmers J, et al. Cloud and aerosol classification for 2.5 years of MAX-DOAS observations in

Wuxi (China) and comparison to independent data sets. Atmos Meas Tech 2015;8(12):5133–56. doi:10.5194/amt-8-5133-2015.

[70] Zhao X, Bognar K, Fioletov V, Pazmino A, Goutail F, Millán L, et al. Assessing the impact of clouds on ground-based UV–visible total column ozone measurements in the high arctic. Atmos Meas Tech 2019;12(4):2463–83. doi:10.5194/amt-12-2463-2019.

[71] Schneider M, Redondas A, Hase F, Guirado C, Blumenstock T, Cuevas E. Comparison of ground-based brewer and FTIR total column O_3 monitoring techniques. Atmos Chem Phys 2008;8(18):5535–50. doi:10.5194/acp-8-5535-2008.