Accuracy, precision, and temperature dependence of Pandora total ozone measurements estimated from a comparison with the Brewer triad in Toronto

Xiaoyi Zhao[1], Vitali Fioletov[2], Kimberly Strong[1], Alexander Cede[3,4], and Jonathan Davies[2]

[1] Department of Physics, University of Toronto, Toronto, M5S 1A7, Canada.

[2] Environment and Climate Change Canada, Toronto, M3H 5T4, Canada.

[3] NASA Goddard Space Flight Center, Greenbelt, MD 20771, USA.

[4] LuftBlick, Kreith, Austria.

Correspondence to: Xiaoyi Zhao (xizhao@atmosp.physics.utoronto.ca) and Vitali Fioletov (vitali.fioletov@canada.ca or vitali.fioletov@outlook.com)

Abstract. This study evaluates the performance of the recently developed Pandora spectrometer by comparing it with the Brewer reference triad. This triad was established by Environment and Climate Change Canada (ECCC) in the 1980s and is used to calibrate Brewer instruments around the world, ensuring high quality total column ozone (TCO) measurements. To reduce stray light, the double Brewer instrument was introduced in 1992, and a new reference triad of double Brewers is also operational at Toronto. Since 2013, ECCC has deployed two Pandora spectrometers co-located with the old and new Brewer triads, making it possible to study the performance of three generations of ozone-monitoring instruments. The statistical analysis of TCO records from these instruments indicates that the random uncertainty for the Brewer is below 0.6 %, while that for the Pandora is below 0.4 %. However, there is a 1 % seasonal difference and a 3 % bias between the standard Pandora and Brewer TCO data, which is related to the temperature dependence and difference in ozone cross sections. A statistical model was developed to remove this seasonal difference and bias. It was based on daily temperature profiles from the European Centre for Medium-Range Weather Forecasts Interim data over Toronto and TCO from the Brewer reference triads. When the statistical model was used to correct Pandora data, the seasonal difference was reduced to 0.25 % and the bias was reduced to 0.04 %. Pandora instruments were also found to have low airmass dependence up to 81.6° solar zenith angle, comparable to double Brewer instruments.

1 Introduction

Routine total column ozone (TCO) measurements started in the 1920s with the Dobson instrument (Dobson, 1968). During the International Geophysical Year, 1957, the worldwide Dobson ozone-monitoring network was formed. Stratospheric ozone has been an important scientific topic since the 1970s and became a matter of intense interest with the discovery and

5   subsequent studies of the Antarctic ozone hole (Farman et al., 1985; Solomon et al., 1986; Stolarski et al., 1986) and depletion on the global scale (Stolarski et al., 1991; Ramaswamy et al., 1992). To improve the accuracy and to automate the TCO measurements, the Brewer spectrophotometer was developed in the early 1980s (Kerr et al., 1981; 1988). In 1988, the Brewer was designated (in addition to the Dobson) as the World Meteorological Organization (WMO) Global Atmosphere Watch (GAW) standard for total column ozone measurement. By 2014, there were more than 220 Brewer instruments

10   installed around the world, with most in operation today. To maintain the measurement stability and characterize each individual Brewer, field instruments need to be regularly calibrated against the traveling standard reference instrument. The traveling standard itself is calibrated against the set of three Brewer instruments (serial numbers 8, 14, and 15) operated by Environment and Climate Change Canada (ECCC), located in Toronto, and known as the Brewer reference Triad (BrT) (Fioletov et al., 2005). Due to the well-known stray light issue in the UV region (Bais et al., 1996; Fioletov et al., 2000), the

15   MkIII Brewer (double Brewer) was introduced in 1992. The double Brewer has two spectrometers in series, significantly improving UV response and measuring global UV spectral irradiance, O_3, SO_2 and aerosol optical depth. The double Brewer instruments also have a set of three instruments (serial numbers 145, 187, and 191) co-located with BrT to form the Brewer reference Triad-Double (BrT-D). Individual Brewer instruments of the BrT and BrT-D are independently calibrated at Mauna Loa, Hawaii every 2-6 years (Fioletov et al., 2005).

The Pandora system was developed at NASA's Goddard Space Flight Center and first deployed in the field in 2006. Pandora instruments are based on a commercial spectrometer with stability and stray light characteristics that make them suitable candidates for both direct-sun and zenith-sky measurements of total column ozone and other trace gases (Herman et al., 2009; Tzortziou et al., 2012). Pandora instruments have been tested and deployed in multiple scientific measurement

25   campaigns around the world. These include the Cabauw Intercomparison Campaign of Nitrogen Dioxide measuring Instruments (CINDI) in the Netherlands in 2009 (Roscoe et al., 2010) and four NASA DISCOVER-AQ campaigns since 2011 (Tzortziou et al., 2012). The Pandora instruments have been used for validation of satellite ozone (Tzortziou et al., 2012) and NO2 (Herman et al., 2009; Tzortziou et al., 2012) measurements. By 2015, several long-term Pandora sites had been established in the United States and worldwide (including Austria, Canada, Canary Islands, Finland, and New Zealand).

30   In 2013, two Pandora instruments (serial number 103 and 104) were deployed at Toronto co-located with BrT and BrT-D on the roof of the ECCC Downsview building (43.782° N, 79.47° W).

The instrument random uncertainties of BrT were analyzed by Kerr et al. (1998) and Fioletov et al. (2005) using similar methods. These methods both require knowledge of the extra-terrestrial calibration (ETC) values, the ozone absorption coefficients, and the Rayleigh scattering coefficients for each instrument. Fioletov et al. (2005) reported that the random uncertainties of individual observations from the BrT are within ±1 % in about 90 % of all measurements. This work takes a

5   different approach, using a statistical variable estimation method to determine the random uncertainties for BrT, BrT-D, and the two Pandora instruments together. The variable estimation method follows the work of Fioletov et al. (2006) to estimate the random uncertainties with the assumption that there is no multiplicative bias between Pandoras and Brewers. Details of the method are provided in Sect. 3.1. Since the instrument random uncertainties for BrT were last reported 10 years ago using data to 2004 (Fioletov et al., 2005), this work provides a new assessment of the performance of both the BrT and BrT-

10   D in recent years, along with a comparison between coincident Brewer and Pandora measurements.

It is well known that the Dobson and Brewer ozone retrievals exhibit dependence on stratospheric temperature (Kerr et al., 1988; Redondas et al., 2014; Scarnato et al., 2009). This is because the retrievals use different wavelengths and ozone cross sections measured at fixed temperatures. Brewer instruments have a very low temperature dependence (typically < 0.1 % K^{-1})

15   ^{1}) (Kerr et al., 1988; Kerr, 2002; Van Roozendael et al., 1998; Scarnato et al., 2009; Herman et al., 2015). For example, Kerr et al. (1988) reported a 0.07 % K^{-1} temperature dependence for Brewer #8 (one of the BrT) and Kerr (2002) reported a 0.094 % K^{-1} temperature dependence for Brewer #14 (one of the BrT). In addition, Scarnato et al. (2009) reported that Brewer instruments (#40, #72, and #156) exhibited less temperature dependence than Dobson instruments (#83 and #101). Redondas et al. (2014) reported a 0.133 % K^{-1} temperature dependence for Dobson #83.

The Pandora ozone retrievals are more sensitive to stratospheric temperatures. In Herman et al. (2015), the temperature dependence for Pandora #34 (0.333 % K^{-1}) was determined by applying retrievals at a series of different ozone temperatures from 215 to 240 K for the ozone cross sections, and then obtaining a linear fit to the percent change. As the small Brewer temperature dependence is known, we use coincident measurements from the BrT and BrT-D to determine the temperature

25   dependence factors for Pandora #103 and #104, and then apply the correction to remove the difference between Pandora and Brewer instruments.

2 Instruments and Datasets

2.1 Pandora

The Pandora spectrometer system uses a temperature-stabilized (1°C) symmetric Czerny-Turner system with a 50 micron

30   entrance slit, and 1200 lines/mm grating. Unlike the Brewer instruments, which only measure intensities at selected wavelengths, the Pandora instruments, with a 2048×64 back-thinned Hamamatsu CCD detector, record spectra from 280 to 530 nm at 0.6 nm resolution (Herman et al., 2015). The spectra are analyzed using the Differential Optical Absorption

Spectroscopy (DOAS) technique (Noxon, 1975; Platt and Stutz, 2008; Solomon et al., 1987; Platt, 1994), in which absorption cross sections for multiple atmospheric absorbers (including ozone, NO_2, SO_2, HCHO, and BrO) are fitted to the spectra (Tzortziou et al., 2012). The Daumont, Brion, and Malicet (DBM) (Daumont et al., 1992; Brion et al., 1993), 1998 ozone cross section at an effective temperature of 225° K is used in the Pandora retrievals (Herman et al., 2015). Additional information on Pandora calibrations and operation can be found in Herman et al. (2015).

Two commercial Pandoras (#103 and #104) were used in this study with no modifications to operational and processing algorithms (available from SciGlob http://www.sciglob.com/). Pandoras #103 and #104 were deployed in Toronto in September 2013, and in this work, all available Pandora data from these instruments are used. Pandora #104 was moved to the Canadian oil sands region in August 2014. Following the work of Tzortziou et al. (2012), the Pandora ozone dataset is filtered to remove data from which the normalized root-mean square (RMS) of weighted spectral fitting residuals is greater than 0.05 and the Pandora calculated standard uncertainty (Tzortziou et al., 2012) in TCO is greater than 2 DU.

2.2 Brewer

The Brewer instruments use a holographic grating in combination with a slit mask to select six channels in the UV (303.2, 306.3, 310.1, 313.5, 316.8, and 320 nm) to be detected by a photomultiplier. The first and second wavelengths are used for internal calibration and measuring SO_2 respectively. The four longer wavelengths are used for the ozone retrieval. The total column of ozone is calculated by analyzing the relative intensities at these different wavelengths using the Bass and Paur (1985) ozone cross sections at a fixed effective temperature of 228.3° K (Kerr, 2002).

Most of the instruments in the BrT (#8, #14, and #15) and BrT-D (#145, #187, and #191) have been in operation since Pandora instruments were deployed. However, there are a few measurement gaps for some of the Brewers. For example, Brewers #14 and #15 were recalibrated at Mauna Loa, Hawaii in October 2013, and Brewer #145 was in Spain in March 2014. We also had to exclude some periods due to instrument malfunction and repairs. The coincident measurement periods for the instruments are shown in Table 1. The data from Brewer and Pandora instruments are both time binned (3 min) for the comparison. Following the work of Tzortziou et al. (2012), the Brewer dataset is filtered to remove data with calculated standard uncertainty in TCO greater than 2 DU. In addition, the Brewer dataset is filtered for clouds by removing data for which the logarithm of the signal at 320 nm is less than the mean value minus two standard deviations (4 % of data was removed with this filter).

2.3 OMI

The Ozone Monitoring Instrument (OMI) is a nadir-viewing near-UV/Vis spectrometer aboard NASA's Earth Observing System (EOS) Aura satellite (launched in July 2004). The OMI instrument measures the solar radiation backscattered by the Earth's atmosphere and surface between 270-500 nm with a spectral resolution of about 0.5 nm (Levelt et al., 2006). The

OMI TCO data are retrieved using both the Total Ozone Mapping Spectrometer (TOMS) technique (developed by NASA (Bhartia and Wellemeyer, 2002) and based on a retrieval using four wavelengths at 313, 318, 331, and 360 nm) and the DOAS technique (developed by KNMI (Veefkind et al., 2006; Kroon et al., 2008) and based on the spectrum measured in the wavelength range 331.1-336.6 nm). The OMI TCO validation done by Balis et al. (2007) shows a globally averaged

5   agreement of better than 1 % for OMI-TOMS data and better than 2 % for OMI-DOAS data in comparison with Brewer and Dobson measurements.

The OMI TCO products used in the present study are the Level-3 Aura/OMI daily global TCO gridded product (OMTO3e) retrieved by the enhanced TOMS Version 8 algorithm (Balis et al., 2007). The OMTO3e data (Bhartia, 2012) are generated

10   by the NASA OMI science team by selecting the best pixel (shortest path length) data from the good quality Level-2 TCO orbital swath data (for example, L2 observations with SZA < 70°; details can be found in (Bhartia, 2012) that fall in the 0.25×0.25° global grids. The OMTO3e data that from the grid point over the ground-based site are used in this work to validate our correction method for Pandora TCO data.

2.4 ECMWF Interim data

15   In this work, the ozone-weighted effective temperature was used to assess the temperature sensitivity of Pandora ozone retrievals. Temperature and ozone profiles were extracted from the European Centre for Medium-Range Weather Forecasts (ECMWF) Interim data for 2013-2015 (Dee et al., 2011) with 0.5°×0.5° spatial resolution on 37 standard pressure levels, available from http://apps.ecmwf.int/datasets/. The ozone-weighted effective temperature (T_{eff}) is calculated based on daily ozone and temperature profiles (at 18:00 UTC) over Toronto, defined as

$$T_{eff} = \sum_{i=6}^{30} w_{eff,i} \cdot T_i \tag{1}$$

$$w_{eff,i} = \frac{n_i}{\sum_{j=6}^{30} n_j} = \frac{MMR_i \cdot p_i / T_i}{\sum_{j=6}^{30} MMR_j \cdot p_j / T_j} \tag{2}$$

where w_{eff} is the weighting function, T_i is the temperature, n_i is the ozone number density, MMR_i is the ozone mass mixing

25   ratio, and p_i is the pressure at pressure level i. In this work, profile data on ECMWF standard pressure levels from #6 to #30 (10-800 mbar) were used to decrease the noise from variable surface temperatures.

3 Statistical Uncertainty Estimation

Figure 1 shows the time series of the total column ozone datasets used in this work. The seasonal cycles of TCO from the ground-based and satellite instruments track each other well, and the high-frequency daily variations from all ground-based

30   instruments are consistent.

By comparing the same quantity retrieved from different remote sensing instruments, we can characterize the differences between them, which are a combination of random uncertainties and systematic bias. Theoretically, information about the random uncertainties can be derived from the measurements themselves (Grubbs, 1948; Toohey and Strong, 2007). The

5   following method for doing this is described in Fioletov et al. (2006), and briefly explained below.

3.1 Method

We model the two types of measured TCO (denoted as M_B and M_P, for Brewer and Pandora, respectively) as simple linear functions of the true TCO value (X) and instrument random uncertainties (δ_B and δ_P), and assume that there is no multiplicative or additive bias between Pandora and Brewer, giving

$$M_B = X + \delta_B$$
$$M_P = X + \delta_P \ . \tag{3}$$

If we assume that the instrument random uncertainties are independent of the measured TCO, the variance of M is the sum of

15   the variances of X and δ,

$$\sigma_{M_B}^2 = \sigma_X^2 + \sigma_{\delta_B}^2$$
$$\sigma_{M_P}^2 = \sigma_X^2 + \sigma_{\delta_P}^2 \ . \tag{4}$$

20   If the difference between Pandora and Brewer does not depend on X (no multiplicative bias), and the random uncertainties of the two instruments are not correlated, then the variance of the difference is equal to the sum of the variance of the random uncertainties,

$$\sigma_{M_B-M_P}^2 = \sigma_{\delta_B}^2 + \sigma_{\delta_P}^2 \ . \tag{5}$$

Since we have the measured TCO and the difference between the Pandora and Brewer datasets, the variance of the TCO and instrument random uncertainties can be solved by

$$\sigma_X^2 = \left(\sigma_{M_B}^2 + \sigma_{M_P}^2 - \sigma_{M_B-M_P}^2\right)/2$$
30    $$\sigma_{\delta_B}^2 = \left(\sigma_{M_B}^2 - \sigma_{M_P}^2 + \sigma_{M_B-M_P}^2\right)/2$$
$$\sigma_{\delta_P}^2 = \left(\sigma_{M_P}^2 - \sigma_{M_B}^2 + \sigma_{M_B-M_P}^2\right)/2 \ . \tag{6}$$

Equation (6) can be used to estimate the standard deviation (SD) of instrument random uncertainties (σ_{δ_B} and σ_{δ_P}) and the SD of ozone variability (σ_X). We do not actually know the variances $\sigma^2_{M_B}$, $\sigma^2_{M_P}$, and $\sigma^2_{M_B-M_P}$; we can only estimate them, with some uncertainty, from the available measurements. It can be shown that the uncertainties in the σ^2_X, $\sigma^2_{\delta_B}$, and $\sigma^2_{\delta_P}$ estimates depend on the sum of all three variances $\sigma^2_{M_B}$, $\sigma^2_{M_P}$, and $\sigma^2_{M_B-M_P}$, and can be high even if the estimated variance itself is low (but one or more of the variances $\sigma^2_{M_B}$, $\sigma^2_{M_P}$, and $\sigma^2_{M_B-M_P}$ are high). The estimates are thus only as accurate as the least accurate of these parameters. The variance estimates can be improved by increasing the number of data points or by reducing variances of X by removing some of the variability. To remove the variability in X, the residual ozone here is defined as the difference between the high-frequency TCO and the low-frequency TCO measured by an instrument,

$$TCO_{res}(t) = TCO_{high-f}(t) - TCO_{low-f}(t) \tag{7}$$

where t is the time of the measurement. For example, the Brewer residual ozone could be the Brewer TCO measurements minus the Brewer ozone daily mean, whereas the corresponding Pandora residual ozone would be the Pandora TCO measurements minus the Pandora ozone daily mean. By subtracting the low-frequency signal, we remove most of the ozone variability. In addition, as proposed in Fioletov et al. (2005), to improve the removal of the bias, we can use the following statistical model to calculate the low-frequency signal:

$$TCO_{low-f}(t) = A_B \cdot I_B + A_P \cdot I_P + B \cdot (t - t_0) + C \cdot (t - t_0)^2 \tag{8}$$

where t is the time of the measurement and t_0 is the time of local solar noon. I_B is an indicator function for the Brewer instrument; it is set to 1 if the TCO is measured by the Brewer and to 0 otherwise. I_P is the indicator function for the Pandora. The coefficients A_B, A_P, B, and C are estimated by the least-squares method for each day. In the following, we will refer to the residual ozone calculated by subtracting the daily mean value as residual type 1, and that obtained by subtracting this 2nd order function as residual type 2.

3.2 Results

In this work, we calculate two different types of residual ozone as defined in Sect. 3.1, and then use them to calculate the instrument random uncertainty with the statistical variable estimation method. The residual types and relevant terminologies are summarized in 2.

Figure 2 shows the Brewer estimated random uncertainties obtained using the two types of residual ozone data (Fig. 2a for residual type 1, Fig. 2b for type 2). For example, in Fig. 2a, the estimated random uncertainty for Brewer #8 using Pandora #103 data (residual type 1, derived from M_{P103}) is shown as a black square in the column for Brewer #8, while its estimated

random uncertainty using Pandora #104 data (residual type 1, derived from M_{P104}) is shown as a red triangle in the same column. Figure 2 demonstrates that type 1 (Fig. 2a) and type 2 (Fig. 2b) residual ozone data provide comparable results, and confirm that Brewer instruments have random uncertainties of 1-2 DU.

5   Figure 2 also shows the Pandora estimated random uncertainties using the two types of residual ozone data (Fig. 2c for residual type 1, Fig. 2d for type 2). For example, in Fig. 2c, the estimated random uncertainty for Pandora #103 using Brewer #8 data is shown as a black square in the column of Brewer #8, while its estimated random uncertainties using other Brewer data are shown by respective Brewer columns. Figure 2 demonstrates that the Pandora instruments have estimated random uncertainties less than 1.5 DU. Slight differences in the estimated Pandora random uncertainties were found using

10   different Brewer instruments. This is due to the sample size; when the sample size is large (> 1200 coincident points, see Table 1), the Pandora estimated random uncertainties from different instruments are more consistent. For example, in Fig. 2c, one of the estimated random uncertainties for Pandora #103 (black square in Brewer #187 column) is below 0.5 DU. This result is undesirable (the value is ~0.5 DU lower than the other values), but not unusual. Dunn (2009) describes this issue in detail and points out that the low (even negative in some cases) variance estimate is due to small sample size. In general,

15   Dunn (2009) concludes that, even with the correct model, the comparisons and estimation of precision are only viable with large sample sizes. Figure 3c shows that the low variance was indeed from the smallest sample size (608 coincident points for Pandora #103 vs. Brewer #187 and 397 for Pandora #104 vs. Brewer #187). In addition, when using the data from the same pair of Brewer and Pandora instruments, the estimated random uncertainty for Pandora is consistently lower than that for Brewer by ~0.5 DU.

Fioletov et al. (2006) estimated natural ozone variability (σ_X) using Eq. (6). However, because we are using the residual ozone instead of the TCO in the statistical analysis, the σ_X calculated from our method is not the estimated natural ozone variability but the estimated residual ozone variability. It can be used to characterize the difference between residual types 1 and 2. Figure 3a shows the estimated residual ozone variability using residual type 1 data, while Fig. 3b shows the variability

25   using residual type 2. Figure 3a and 3b demonstrate that residual type 1 data has larger variability than type 2 data, indicating that using the daily mean value as the low-frequency signal did not fully remove the natural ozone variability. Ideally, the random uncertainty estimate should only contain random noise caused by the instrument and no natural ozone variation. Scatter plots of Brewer vs. Pandora residual ozone (Fig. 4) illustrate the same results. Figure 4 shows that the correlation coefficients for residual type 1 (R = 0.813 for Brewer #8 vs. Pandora #103, see Fig. 4a; 0.909 for Brewer #8 vs.

30   Pandora #104, see Fig. 4b) are higher than the ones for residual type 2 (0.333 for Brewer #8 vs. Pandora #103, see Fig. 4c; 0.688 for Brewer #8 vs. Pandora #104, see Fig. 4d). The low correlation coefficients for ozone residual type 2 data indicate that the ozone variability has been largely removed from Pandora and Brewer data. Thus when we use residual ozone type 2, even with relatively small sample size, the estimated uncertainties for Pandoras are still consistent with those obtained from comparisons with other Brewers having larger sample sizes (see Fig. 2c and 2d, Brewer #187 column).

To summarize, we tested two different methods for calculating residual ozone, and applied them in the statistical uncertainty estimation. In general, Fig. 2 demonstrates that the Pandora TCO data has ~0.5 DU smaller estimated random uncertainties compared to the Brewer TCO data. The mean estimated random uncertainties for BrT and BrT-D are in the range of 1-2 DU (~0.6 %). The mean estimated random uncertainties for Pandora #103 and #104 are in the range of 0.5-1.5 DU (~0.4 %). These results confirm the quality of the TCO data, with all eight instruments meeting the GAW requirement for a precision better than 1 % to measure ozone (WMO, 2014).

4 Temperature Dependence Effect and Correction

4.1 Method

When comparing Pandora and Brewer TCO data, we can see a clear seasonal structure and a bias in the difference and ratio. Figure 5a shows the time series of Brewer #14 – Pandora #103 TCO difference; the seasonal amplitude is 3-4 DU and the mean bias is 10.81 DU. Figure 5b (which uses the corrected data) will be discussed in Sect. 4.2. The Lowess(x) fit (the dashed line) is based on local least squares fitting applied to a specified x fraction of the data (Cleveland and Devlin, 1988). The bias between Pandora and Brewer TCO is mainly due to the fact that both retrievals depend on the choice of ozone absorption cross section (Scarnato et al., 2009; Herman et al., 2015). The Brewer TCO in this work was retrieved using the standard Brewer network operational ozone cross section (Bass and Paur, 1985), while the Pandora TCO was retrieved using the standard Pandora network operational ozone cross section (the DBM ozone cross section). Redondas et al. (2014) reported that by changing the Brewer operational ozone cross section from Bass and Paur (1985) to that of Daumont et al. (1992) (DBM) will change the calculated TCO by -3.2 %. In addition to the offset caused by the use of different ozone cross sections, the seasonal difference between Pandora and Brewer TCO data is due to their differing temperature dependence, which varies from instrument to instrument because of the differences in ozone retrieval algorithm and instrument design. Moreover, even for the same type of instrument, the temperature sensitivity can be different due to imperfections in the wavelength settings and slit function for each individual instrument. We will study these differences (offset and temperature effect) by using the standard TCO products from Pandora and Brewer instruments.

In this work, we use ECMWF Interim ozone and temperature profiles to calculate daily ozone effective temperature (described in Sect. 2.4). Then we use the following simple linear regression model to find the temperature dependence factor for Pandora instruments,

$$\frac{M_B}{M_P} = a \cdot \left(T_{eff} - 225 \right) + b \tag{9}$$

where a is the temperature dependence factor for Pandora, b is the (systematic) multiplicative bias between Pandora and Brewer, and 225 refers to effective temperature of 225° K for ozone cross-sections used in the Pandora retrievals. Here, the M_B and M_P are TCO daily means measured by the Brewer and Pandora respectively. To increase the number of coincident data points, the M_B dataset is formed by merging all measurements from the six Brewers (see Table 1). A successfully

5   merged M_B data point has coincident measurements from at least two Brewers, to avoid domination by a single instrument. The coincident time period of the M_B and M_{P103} datasets is from October 2013 to December 2015 with 272 coincident days (points). Figure 6 shows the linear regression results for Pandoras #103 and #104. We found the relative temperature dependence factor for Pandora #103 to be 0.247 ± 0.013 % K^{-1} (from the term a in Eq. (9)), with a 2.2 ± 0.1 % multiplicative bias (from the term b in Eq. (9)). Although Pandora #104 only has measurements from January to April 2014 (53 coincident

10   days), the linear regression still results in a similar temperature dependence factor (0.255 ± 0.040 % K^{-1}) and the same bias as Pandora #103. The correlation coefficients for those two linear regressions are 0.91 and 0.89 respectively.

We applied the Pandora temperature dependence factors to the Pandora TCO to remove its bias and seasonal difference relative to Brewer TCO data. Similar to the correction function used in Herman et al. (2015) for Pandora #34, we used the

15   following function to correct Pandora TCO data:

$$M_{corr} = M_P \cdot \left(a \cdot \left(T_{eff} - 225 \right) + b \right) \qquad (10)$$

where M_{corr} is corrected Pandora TCO, and other terms are as defined for Eq. (9). For the Pandora #103 dataset, this becomes

$$M_{corr} = M_{P103} \cdot \left(0.00247 \cdot \left(T_{eff} - 225 \right) + 1.022 \right) \qquad (11)$$

where M_{p103} is the TCO data from Pandora #103. The temperature dependence factor (0.247 ± 0.013 % K^{-1}) and the multiplicative bias (1.022) are found in Fig. 6. The same regression model and method give 0.255 ± 0.040 % K^{-1} temperature

25   dependence factor with a 2 % multiplicative bias to Pandora #104, and hence

$$M_{corr} = M_{P104} \cdot \left(0.00255 \cdot \left(T_{eff} - 225 \right) + 1.022 \right) \qquad (12)$$

where M_{p104} is the Pandora #104 TCO. For comparison, Herman et al. (2015) derived the correction function for Pandora #34

30   as

$$M_{corr} = M_{P34} \cdot \left(0.00333 \cdot \left(T_{eff} - 225 \right) + 1 \right) \qquad (13)$$

where the 0.00333 (0.333 % K^{-1}) is the temperature dependence factor for Pandora #34. Note that this value was determined by applying retrievals using ozone cross sections from 215 to 240 K, and then obtaining a linear fit to the percent change (Herman et al., 2015). However in this work, the factors for Pandora #103 and #104 were found by statistical analysis (comparison) of the Pandora and Brewer TCO datasets. Thus our temperature dependence factor combines the temperature

5 sensitivity from both Pandora and Brewer instruments, and describes the relative temperature sensitivity between the Pandora and Brewer standard TCO products. We call it a "relative temperature dependence factor" (RTDF), while that from Herman et al. (2015) is an "absolute temperature dependence factor" (ATDF). Although the RTDF is a non-linear combination of ATDF from both Pandora and Brewer (note that the Pandora used an ozone cross section at an effective temperature of 225 K, while the Brewer used that at 223.8 K), we can still make a simple linear estimation of the RTDF from

10 reported ATDFs. In fact, the reported ATDF for Pandora #34 (0.333 % K^{-1}, (Herman et al., 2015)) minus the reported ATDF for Brewer #8 and #14 (0.07 and 0.094 % K^{-1}, (Kerr et al., 1988), 2002) gives relative numbers (0.26 and 0.24 % K^{-1}) that are close to our model-calculated RTDF (~0.25 % K^{-1}). In our correction functions (Eqs. (11-12)), we have a constant b term of 1.022 given 0.001 uncertainty, which indicates a multiplicative bias of ~2 % (not caused by the temperature effect) between the Pandora and Brewer instruments due to their different selection of ozone cross sections.

Merging data from all six Brewers could lead to variation of the Brewer temperature dependence, so we performed sensitivity tests on the dataset. Figure 3 summarizes the tests; the combined Brewer data are merged from all available Brewer data during the data period indicated in the table. Figure 7 shows the RTDFs, multiplicative bias, correlation coefficient, and number of data points for the thirteen sensitivity tests. Tests 1 and 2 are the results adapted from Fig. 6. Due

20 to the small data size, the RTDF for test 2 has larger error bars than test 1. Test 3 shows Pandora #103 RTDF using combined Brewer data for the same time period as Pandora #104. Pandora #103 has a measurement gap from Aug. to Dec. 2014 due to instrument failure (see Fig. 1), hence, tests 4 and 5 use combined Brewer data for 2013-2014 (~ one year coverage, before the instrument failure of Pandora #103) and 2015 (one year coverage, after Pandora #103 was repaired) separately. Brewer #191 was one of the most reliable Brewer instruments during the comparison period. Thus tests 6-8 use

25 only Brewer #191 data; test 6 uses all available data (2013-2015), test 7 uses only 2013-2014 data (before the instrument failure of Pandora #103), test 8 uses 2015 data (after Pandora #103 was repaired). Tests 9-13 use individual Brewer data (all available data for each individual Brewer). For the thirteen tests, the RTDFs (see Fig. 7a) are in the range of 0.24-0.29 %, and the multiplicative biases (see Fig. 7b) are in the range of 1.7-2.5 %. The correlation coefficients (see Fig. 7c) for most tests are above 0.8. In general, the RTDFs found for the Pandora instruments are stable when derived from combined Brewer

30 data or reliable individual Brewer data.

4.2 Results

4.2.1 Pandora TCO Correction

As an example, Fig. 5 shows the time series of Brewer #14 – Pandora #103 TCO differences, before and after applying the Pandora correction (Eq. (11)). A clear seasonal signal is seen due to the variation of T_{eff} before we apply the temperature

5 dependence correction (see Fig. 5a). Figure 8 shows scatter plots of Pandora #103 versus Brewer #14 TCO. In Fig. 8a, the linear regression (green line, weighted accounting for uncertainties from both measurements (York et al., 2004)) between Pandora #103 and Brewer #14 gives a slope of 1.023, an offset of -18.486 DU, and strong correlation (R = 0.9954). Forcing the intercept to zero gives a slope of 0.969, indicating -3.1 % mean bias. This is consistent with the work of Redondas et al. (2014), which showed that changing the Brewer ozone cross section from Bass and Paur to DBM changed the Brewer TCO

10 by -3.2 %. By colour coding the scatter points, it is obvious that this non-ideal slope and offset are related to T_{eff}. After applying the correction, the seasonal Brewer – Pandora difference disappears as seen in Fig. 5b, and the linear regression (green line) gives a slope of 1.008, an offset of -2.678 DU, and an improved correlation (R = 0.9982) (see Fig. 8b). Linear fitting with zero intercept gives a slope of 1.001, indicating that the correction improves the mean bias between Pandora and Brewer TCO from -3.1 % to 0.1 %.

To calculate the effective temperature, we use daily temperature and ozone profiles from ECMWF Interim data at 18:00 UTC for Toronto, but Herman et al. (2015) used monthly averaged climatology data (see ftp://toms.gsfc.nasa.gov/ pub/ML_climatology) for latitudes of 30-40º N and 40-50º N to form an average suitable for Boulder (40º N). To understand the difference due to the selection of T_{eff}, we adapted the climatology data used in Herman et al. (2015), and used the data

20 from 40-50º N to calculate effective ozone temperature for Toronto (44º N). Figure 9 shows the comparison between the ECMWF daily T_{eff} and the NASA monthly climatology T_{eff}. A sudden cooling event happened at Toronto on 29-30 January 2014, for which the difference between the daily and monthly T_{eff} was -10 K. Figure 10 shows the time series of TCO difference (combined Brewer − Pandora #103) before and after applying the temperature dependence correction using both the monthly climatology T_{eff} and daily T_{eff}. Because the monthly climatology T_{eff} does not reflect the low temperature during

25 those two days, the correction function (see Eq. (11)) overcompensated for the temperature effect (the minimum delta ozone value on 29 January changed from -8 DU in Fig. 10a to -14 DU in Fig. 10b). The low-temperature event was captured by the daily T_{eff}, thus the compensation from the temperature effect was reasonably small when using ECMWF daily T_{eff} (the minimum value was -7 DU, see Fig. 10c). In general, the ECMWF daily T_{eff} can better capture some ozone variation events that are associated with rapid temperature changes.

Figure 11 shows time series of the monthly average TCO difference in percentage before and after applying the temperature dependence correction for eight pairs of instruments (six individual Brewers vs. Pandora #103, combined Brewer vs. Pandora #103, and combined Brewer vs. Pandora #104). Figure 11a shows that both Pandora #103 and #104 have similar

offsets relative to the Brewers before applying the correction to Pandora data. In addition, the seasonal variations are consistent when comparing Pandora #103 to six individual Brewers (see Fig. 11a). After applying the TCO corrections (Fig. 11b), the seasonal differences decreased from ±1.02 % to ±0.25 % for Pandora #103 and from ±0.40 % to ±0.25 % for Pandora #104, as did the offset which decreased from 2.92 % to -0.04 % for Pandora #103 and from 2.11 % to -0.01 % for

5   Pandora #104. The 1σ uncertainty in Fig. 11b shows that, statistically, the corrected Pandora datasets have no significant seasonal differences or offsets compared to the Brewer datasets.

4.2.2 Comparison with OMI

To further validate the temperature dependence correction for the Pandora data, we used OMI ozone data (version OMTO3e). Pandora data are averaged within ±10 min of OMI overpass times. In Fig. 12, scatter plots of OMI vs. Pandora

10   TCO are shown in panels a and b; OMI vs. corrected Pandora TCO (using Eq. (11) and (12) with the correction functions found from our statistical model) is shown in panels c and d; OMI vs. corrected Pandora TCO (using Eq. (13) with the correction function from (Herman et al., 2015)) is shown in panels e and f. All the Pandora TCO corrections shown in Fig. 12 used the same T_{eff} calculated with the ECMWF Interim daily ozone data.

15   Figure 12a and c show that, after applying the TCO correction (Eq. (11)) to Pandora #103, the slope of the linear regression improved from 0.987 to 0.990, the offset improved from 14.84 to -3.59 DU, the correlation coefficient improved from 0.987 to 0.991, and the mean bias between OMI and Pandora improved from 3.1 % to 0.02 %. Similar improvement is seen in the comparison between Pandora #104 and OMI (see Fig. 12b and d), although the size of the coincident measurement dataset is smaller, with the mean bias improving from 1.5 % to -0.6 %. In addition, Fig. 12e and f show that, by using the correction

20   function from (Herman et al., 2015), the comparisons also improve, although 1.9 % (1.4 %) bias remains for Pandora #103 (#104) (indicated by the slop of linear fit with force the intercept to zero, see the green lines in Fig. 12). Note that the ATDF in Herman et al. (2015) is only 0.08 % K^{-1} higher than our RTDF.

Figure 13a and b show the monthly mean time series of the OMI – Pandora TCO percentage difference, before and after

25   applying the three correction functions. All three correction models reduced the difference between Pandora and OMI. Our relative correction model (Eq. (11) and (12)) reduces the seasonal difference (indicated by the δ of the percentage monthly delta ozone) between Pandora #103 and OMI from ±1.68 % to ±1.00 %, with the mean bias decreasing from 2.65 % to -0.19 % (the mean of the percentage monthly delta ozone). Pandora #104 has a similar improvement. The absolute correction model (Eq. (13)) reduces the seasonal difference between Pandora #103 and OMI to 0.87 %, with the mean bias decreased to

30   1.71 %. The reduction in the mean bias between Pandora and OMI is better for the relative correction model. This result is consistent with Balis et al. (2007) who showed that the global average difference between OMI-TOMS and Brewer instruments is within 0.6 %.

Balis et al. (2007) reported that the time series of globally averaged differences between OMI-TOMS and Brewer instruments shows almost no annual variation, and the OMI-TOMS data theoretically have no temperature dependence (McPeters and Labow, 1996; Bhartia and Wellemeyer, 2002). By using our relative correction, the corrected Pandora TCO should have similar performance to the Brewer TCO. Figure 13c shows the difference between the absolute correction method and the relative correction method. Although both methods removed some of the seasonal signal (reduced from 1.68 % to 1.00 % for the relative correction, and 0.87 % for the absolute correction), Fig. 13c shows that there is still a weak seasonal signal residual (0.39 %) left between these two methods.

5 Stray Light Effect

It is well known that direct-sun UV spectrometers are affected by stray light when the solar zenith angle is too large. In general, when the ozone airmass factor (AMF) is larger than 3 (SZA > 70°), the retrieved TCO will show an unrealistic decrease with increasing SZA (thus this effect is also known as the airmass dependence effect). In general, the stray light from longer wavelengths results in overestimation of the UV signal at short wavelengths and makes the measured UV signal in that part of the spectrum less sensitive to TCO. The double Brewer spectrometer was introduced in 1996, which uses two spectrometers in series to reduce the stray light (Bais et al., 1996; Wardle et al., 1996; Fioletov et al., 2000). The BrT-D has the advantage of very low internal stray light fraction (10^{-7}, stray light signal divided by total signal) compared to BrT (10^{-5}) in the 300-330 nm spectral range (Fioletov et al., 2000; Tzortziou et al., 2012). For Pandora instruments, a UV340 filter is used to remove most of the stray light that originates from wavelengths longer than 380 nm (Herman et al., 2015). A typical UV340 filter has a small leakage (5 %) at ~720 nm, which misses the detector and hits the internal baffles. Further stray light correction is done by subtracting the signal of pixels corresponding to 280 to 285 nm (which contain almost zero direct illumination) from the rest of the spectrum. However, a very small (but unknown) amount of this stray light may scatter on to the detector (Herman et al., 2015). Tzortziou et al. (2012) tested the stray light effect for Pandora #34 and Brewer #171 and concluded that the Pandora stray light fraction (~10^{-5}) was comparable to the single Brewer. Pandora ozone retrievals are accurate up to a slant column between 1400 and 1500 DU or 70 and 80° SZA, depending on the TCO amount (Herman et al., 2015).

In this work, to assess the airmass dependence, we compared Brewer TCO to the corrected Pandora TCO data. Figure 14 shows an example of the Brewer/Pandora ratio as a function of ozone AMF (reported value in Brewer data) before and after applying the TCO correction (Eq. (11)), with the data points grouped by effective temperature. Before applying the correction (Fig. 14a), the linear fits show consistently low (-0.1 to 0.5 %) relative AMF dependence between Brewer and Pandora (defined as the slope of the linear fit) for each T_{eff} group. However, the linear fit to the whole dataset (all effective temperatures, black line) shows that the relative AMF dependence is -0.007. Figure 14b shows that the correction changed the slope of the black line to 0.001; removing the temperature effect for the Pandora dataset thus reduces the relative AMF

dependence from -0.7 % to -0.1 %. To characterize only the airmass dependence, we therefore removed the temperature dependence effect from the Pandora dataset.

To show how the different instrument designs affect the stray light performance, we merged the six Brewer datasets into two groups (BrT and BrT-D) to compare with the corrected Pandora data. Figure 15 shows the (Brewer-Pandora)/Brewer percentage difference as a function of ozone AMF. In Sections 3 and 4, the TCO data with ozone AMF > 3 was discarded. The purpose of this filter was to ensure that only the best direct-sun measurements (with low airmass dependence) from both instruments were used. However, to study the instrument performance for large AMFs, and also to characterize the performance of Brewer and Pandora instruments, we changed the AMF threshold from 3 to 6. Figure 15 indicates that Pandora, BrT, and BrT-D instruments have similar airmass dependence for ozone AMF < 3 (~71° SZA), consistent with the result reported by Tzortziou et al. (2012). Pandora and BrT-D have similar AMF dependence up to ozone AMF of 5.5-6 (80.6-81.6° SZA), but Pandora and BrT diverge above AMF of 3-4 (71-76° SZA). In general, these results indicate the Pandora and BrT-D instruments have very good stray light control.

6 Conclusions

The instrument random uncertainty, TCO temperature dependence, and ozone airmass dependence have been determined using two Pandora and six Brewer instruments. In general, Pandora and Brewer instruments both have very low random uncertainty (< 2 DU) in the total column ozone measurements, with that for Pandora being ~0.5 DU lower than Brewer. This indicates that Pandora instruments could provide more precise measurements than the Brewer for the study of small-scale (temporal and magnitude) atmospheric changes. This work confirms the quality of the TCO data, with all eight instruments meeting the GAW requirement for a precision better than 1% (WMO, 2014), however, the Brewer instruments have smaller ozone temperature dependence than the Pandoras.

By using the ECMWF Interim and Brewer ozone data in the statistical method, we successfully corrected the Pandora TCO to decrease its temperature dependence. We found relative temperature dependence factors of 0.247 % K^{-1} for Pandora #103 and 0.255 % K^{-1} for Pandora #104 against the Brewer instruments. This relative temperature dependence factor is comparable to the absolute temperature dependence factors previously found for Pandora (0.333 % K^{-1}, by applying retrievals with different ozone cross sections, (Herman et al., 2015) and Brewers (0.07-0.094 % K^{-1}, (Kerr et al., 1988; Kerr, 2002). In addition, a 2 % multiplicative bias was found between the Pandora and Brewer standard TCO products, which is due to the different ozone cross sections used in the retrievals. After applying the corrections, the annual seasonal difference between Pandora and Brewer instruments decreased from ±1.02 to ±0.25 % and the mean bias decreased from 2.92 to 0.04 %. In addition to using model ozone data (ECMWF Interim for our case) to calculate the effective ozone temperature, it

could also be estimated from Brewer or Pandora measurements (Kerr, 2002; Tiefengraber et al., 2016), however, at a cost of decreased TCO measurement precision.

This study confirmed that the Pandora and Brewer TCO data have negligible airmass dependence when the ozone AMF < 3.

5   The Pandora and BrT instruments have similar airmass dependence (relative airmass dependence < ±0.1 %) up to 71° SZA (AMF < 3); the Pandora and BrT-D instruments have very good stray light control, and their AMF dependence is comparably low up to 81.6° SZA (within 1 % up to AMF = 5.5 and within 1.5 % up to AMF = 6).

10   *Acknowledgments.* Data from the BrT, BrT-D and Pandora instruments are available through Environment and Climate Change Canada (contact Dr. Vitali Fioletov, vitali.fioletov@canada.ca). The final version of the Brewer data is (or will be) available from the World Ozone and UV Data Centre (www.woudc.org). Any additional data may be obtained from Xiaoyi Zhao (xizhao@atmosp.physics.utoronto.ca). X. Zhao was partially supported by the NSERC CREATE Training Program in Arctic Atmospheric Science. We thank ECMWF for providing the Interim model data, and the NASA OMI ozone retrieval

15   team   for   providing   the   OMTO3e   data   (available   from   the   GES   DISC: http://disc.sci.gsfc.nasa.gov/datacollection/OMTO3e_V003.html).

References

Bais, A., Zerefos, C., and McElroy, C.: Solar UVB measurements with the double-and single-monochromator Brewer ozone spectrophotometers, Geophys. Res. Lett., 23, 833-836, 1996.

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

Bass, A., and Paur, R.: The ultraviolet cross-sections of ozone: I. The measurements, in: Atmospheric Ozone, Springer, 606-610, 1985.

25   Bhartia, P., and Wellemeyer, C.: OMI TOMS-V8 Total O_3 algorithm, algorithm theoretical baseline document: OMI ozone products, NASA God-dard Space Flight Cent, Greenbelt, Md., 2002.

Bhartia, P.: OMI/Aura TOMS-Like Ozone and Radiative Cloud Fraction Daily L3 Global 0.25x0.25 deg, NASA Goddard Space Flight Center, 10.5067/Aura/OMI/DATA3002, 2012.

Brion, J., Chakir, A., Daumont, D., Malicet, J., and Parisse, C.: High-resolution laboratory absorption cross section of O_3.
30   Temperature effect, Chem. Phys. Lett., 213, 610-612, 10.1016/0009-2614(93)89169-I, 1993.

Brion, J., Chakir, A., Charbonnier, J., Daumont, D., Parisse, C., and Malicet, J.: Absorption Spectra Measurements for the Ozone Molecule in the 350–830 nm Region, J. Atmos. Chem., 30, 291-299, 10.1023/a:1006036924364, 1998.

Cleveland, W. S., and Devlin, S. J.: Locally weighted regression: an approach to regression analysis by local fitting, J. Am. Stat. Assoc. , 83, 596-610, 1988.

35   Daumont, D., Brion, J., Charbonnier, J., and Malicet, J.: Ozone UV spectroscopy I: Absorption cross-sections at room temperature, J. Atmos. Chem., 15, 145-155, 10.1007/bf00053756, 1992.

Dee, D. P., Uppala, S. M., Simmons, A. J., Berrisford, P., Poli, P., Kobayashi, S., Andrae, U., Balmaseda, M. A., Balsamo, G., Bauer, P., Bechtold, P., Beljaars, A. C. M., van de Berg, L., Bidlot, J., Bormann, N., Delsol, C., Dragani, R., Fuentes, M., Geer, A. J., Haimberger, L., Healy, S. B., Hersbach, H., Hólm, E. V., Isaksen, L., Kållberg, P., Köhler, M., Matricardi, M.,

McNally, A. P., Monge-Sanz, B. M., Morcrette, J. J., Park, B. K., Peubey, C., de Rosnay, P., Tavolato, C., Thépaut, J. N., and Vitart, F.: The ERA-Interim reanalysis: configuration and performance of the data assimilation system, Q. J. Roy. Meteor. Soc., 137, 553-597, 10.1002/qj.828, 2011.

Dobson, G. M. B.: Exploring the atmosphere, 2nd ed., Clarendon Press, Oxford, xv, 209 p. pp., 1968.

5   Dunn, G.: Statistical evaluation of measurement errors: Design and analysis of reliability studies, John Wiley & Sons, 2009.

Farman, J. C., Gardiner, B. G., and Shanklin, J. D.: Large losses of total ozone in Antarctica reveal seasonal ClO_x/NO_x interaction, Nature, 315, 207-210, 1985.

Fioletov, V., Kerr, J., Wardle, D., and Wu, E.: Correction of stray light for the Brewer single monochromator, Proceedings of the Quadrennial Ozone Symposium, Japan, 2000, 3-8.

10 Fioletov, V., Kerr, J., McElroy, C., Wardle, D., Savastiouk, V., and Grajnar, T.: The Brewer reference triad, Geophys. Res. Lett., 32, 10.1029/2005GL024244., 2005.

Fioletov, V., Tarasick, D., and Petropavlovskikh, I.: Estimating ozone variability and instrument uncertainties from SBUV (/2), ozonesonde, Umkehr, and SAGE II measurements: Short-term variations, J. Geophys. Res., 111, 10.1029/2005jd006340, 2006.

15 Grubbs, F. E.: On estimating precision of measuring instruments and product variability, J. Am. Stat. Assoc. , 43, 243-264, 1948.

Herman, J., Cede, A., Spinei, E., Mount, G., Tzortziou, M., and Abuhassan, N.: NO_2 column amounts from ground-based Pandora and MFDOAS spectrometers using the direct-Sun DOAS technique: Intercomparisons and application to OMI validation, J. Geophys. Res., 114, 10.1029/2009JD011848, 2009.

20 Herman, J., Evans, R., Cede, A., Abuhassan, N., Petropavlovskikh, I., and McConville, G.: Comparison of ozone retrievals from the Pandora spectrometer system and Dobson spectrophotometer in Boulder, Colorado, Atmos. Meas. Tech., 8, 3407-3418, 2015.

Kerr, J., McElroy, C., and Olafson, R.: Measurements of ozone with the Brewer ozone spectrophotometer, Proceedings of Quadrennial International Ozone Symposium, Boulder, CO, 1981, 74-79.

25 Kerr, J., Asbridge, I., and Evans, W.: Intercomparison of total ozone measured by the Brewer and Dobson spectrophotometers at Toronto, J. Geophys. Res., 93, 11129-11140, 1988.

Kerr, J., McElroy, C., and Wardle, D.: The Brewer instrument calibration center 1984-1996, Proceedings of the Quadrennial Ozone Symposium, 1998, 915-918.

Kerr, J.: New methodology for deriving total ozone and other atmospheric variables from Brewer spectrophotometer direct
30 sun spectra, J. Geophys. Res., 107, 10.1029/2001JD001227, 2002.

Kroon, M., Veefkind, J. P., Sneep, M., McPeters, R. D., Bhartia, P. K., and Levelt, P. F.: Comparing OMI-TOMS and OMI-DOAS total ozone column data, J. Geophys. Res., 113, 10.1029/2007jd008798, 2008.

Levelt, P. F., Hilsenrath, E., Leppelmeier, G. W., Van den Oord, G. H., Bhartia, P. K., Tamminen, J., De Haan, J. F., and Veefkind, J. P.: Science objectives of the ozone monitoring instrument, Geos. and Remo. Sens., IEEE Trans. on, 44, 1199-
35 1208, 2006.

McPeters, R. D., and Labow, G. J.: An assessment of the accuracy of 14.5 years of Nimbus 7 TOMS version 7 ozone data by comparison with the Dobson network, Geophys Res Lett, 23, 3695-3698, 10.1029/96gl03539, 1996.

Noxon, J.: Nitrogen dioxide in the stratosphere and troposphere measured by ground-based absorption spectroscopy, Science, 189, 547-549, 1975.

40 Platt, U.: Differential optical absorption spectroscopy (DOAS), Air Monit. Spectro. Tech., 127, 27-84, 1994.

Platt, U., and Stutz, J.: Differential Optical Absorption Spectroscopy: Principles and Applications, Springer, Berlin, 597 pp., 2008.

Ramaswamy, V., Schwarzkopf, M. D., and Shine, K. P.: Radiative forcing of climate from halocarbon-induced global stratospheric ozone loss, Nature, 355, 810-812, 1992.

45 Redondas, A., Evans, R., Stuebi, R., Köhler, U., and Weber, M.: Evaluation of the use of five laboratory-determined ozone absorption cross sections in Brewer and Dobson retrieval algorithms, Atmos. Chem. Phys., 14, 1635-1648, 10.5194/acp-14-1635-2014, 2014.

Roscoe, H. K., Van Roozendael, M., Fayt, C., du Piesanie, A., Abuhassan, N., Adams, C., Akrami, M., Cede, A., Chong, J., Clemer, K., Friess, U., Ojeda, M. G., Goutail, F., Graves, R., Griesfeller, A., Grossmann, K., Hemerijckx, G., Hendrick, F.,
50 Herman, J., Hermans, C., Irie, H., Johnston, P. V., Kanaya, Y., Kreher, K., Leigh, R., Merlaud, A., Mount, G. H., Navarro,

M., Oetjen, H., Pazmino, A., Perez-Camacho, M., Peters, E., Pinardi, G., Puentedura, O., Richter, A., Schonhardt, A., Shaiganfar, R., Spinei, E., Strong, K., Takashima, H., Vlemmix, T., Vrekoussis, M., Wagner, T., Wittrock, F., Yela, M., Yilmaz, S., Boersma, F., Hains, J., Kroon, M., Piters, A., and Kim, Y. J.: Intercomparison of slant column measurements of NO_2 and O_4 by MAX-DOAS and zenith-sky UV and visible spectrometers, Atmos. Meas. Tech., 3, 1629-1646, 10.5194/amt-3-1629-2010, 2010.

Scarnato, B., Staehelin, J., Peter, T., Gröbner, J., and Stübi, R.: Temperature and slant path effects in Dobson and Brewer total ozone measurements, J. Geophys. Res., 114, 10.1029/2009JD012349, 2009.

Solomon, S., Garcia, R. R., Rowland, F. S., and Wuebbles, D. J.: On the depletion of Antarctic ozone, Nature, 321, 755-758, 1986.

Solomon, S., Schmeltekopf, A., and Sanders, R.: On the interpretation of zenith sky absorption measurements, J. Geophys. Res., 2, 8311-8319, 1987.

Stolarski, R. S., Krueger, A. J., Schoeberl, M. R., McPeters, R. D., Newman, P. A., and Alpert, J. C.: Nimbus 7 satellite measurements of the springtime Antarctic ozone decrease, Nature, 322, 808-811, 10.1038/322808a0, 1986.

Stolarski, R. S., Bloomfield, P., McPeters, R. D., and Herman, J. R.: Total Ozone trends deduced from Nimbus 7 Toms data, Geophys. Res. Lett., 18, 1015-1018, 10.1029/91gl01302, 1991.

Tiefengraber, M., Cede, A., and Cede, K.: ESA Ground-Based Air-Quality Spectrometer Validation Network and Uncertainties Study: Report on Feasibility to Retrieve Trace Gases other than O_3 and NO_2 with Pandora, LuftBlick, 32, 2016.

Toohey, M., and Strong, K.: Estimating biases and error variances through the comparison of coincident satellite measurements, J. Geophys. Res., 112, 10.1029/2006JD008192, 2007.

Tzortziou, M., Herman, J. R., Cede, A., and Abuhassan, N.: High precision, absolute total column ozone measurements from the Pandora spectrometer system: Comparisons with data from a Brewer double monochromator and Aura OMI, J. Geophys. Res., 117, 10.1029/2012JD017814, 2012.

Van Roozendael, M., Peeters, P., Roscoe, H., De Backer, H., Jones, A., Bartlett, L., Vaughan, G., Goutail, F., Pommereau, J.-P., and Kyro, E.: Validation of ground-based visible measurements of total ozone by comparison with Dobson and Brewer spectrophotometers, J. Atmos. Chem., 29, 55-83, 1998.

Veefkind, J. P., Haan, J. F. d., Brinksma, E. J., Kroon, M., and Levelt, P. F.: Total ozone from the ozone monitoring instrument (OMI) using the DOAS technique, Geos. and Remo. Sens., IEEE Trans. on, 44, 1239-1244, 10.1109/tgrs.2006.871204, 2006.

Wardle, D., McElroy, C., Kerr, J., Wu, E., and Lamb, K.: Laboratory tests on the double Brewer spectrophotometer, Proceedings of the Quadrennial Ozone Symposium, Italy, 1996, 997-1000.

WMO: Scientific Assessment of Ozone Depletion: 2014, Global Ozone Research and Monitoring Project-Report No. 55, 416, 2014.

York, D., Evensen, N. M., Martınez, M. L., and Delgado, J. D. B.: Unified equations for the slope, intercept, and standard errors of the best straight line, Am. J. Phys., 72, 367-375, 2004.

Table 1: Coincident measurement periods and number of data points for comparisons between Pandora and Brewer instruments.

		Pandora#103	Pandora#104
Brewer#8	Coincident period	18 Oct 2013 to 14 May 2015	20 Jan 2014 to 08 Aug 2014
	Coincident data points	5008	2671
Brewer#14	Coincident period	25 Nov 2013 to 24 Dec 2015	16 Feb 2014 to 08 Aug 2014
	Coincident data points	7797	1701
Brewer#15	Coincident period	31 Nov 2013 to 31 Jul 2014	20 Jan 2014 to 08 Aug 2014
	Coincident data points	2297	1376
Brewer#145	Coincident period	15 Jan 2015 to 24 Dec 2015	N/A
	Coincident data points	1474	N/A
Brewer#187	Coincident period	18 Oct 2013 to 23 Apr 2014	20 Jan 2014 to 23 Apr 2014
	Coincident data points	608	397
Brewer#191	Coincident period	20 Nov 2013 to 24 Dec 2015	21 Jan 2014 to 08 Aug 2014
	Coincident data points	5359	1490

Table 2: Definition of terminologies used in the uncertainty estimation.

	Definition
Estimated random uncertainty (σ_δ)	Random uncertainty estimated using the statistical variable estimation method described in Sect. 3.1
$TCO_{high\text{-}f}$	High-frequency TCO measurements, averaged in 3 min bin
$TCO_{low\text{-}f\ (daily\text{-}mean)}$	Low-frequency TCO, calculated as the daily mean TCO
$TCO_{low\text{-}f\ (2nd\ order\ function)}$	Low-frequency TCO, calculated using the 2nd order function (Eq. (8))
Residual type 1	$TCO_{high\text{-}f} - TCO_{low\text{-}f\ (daily\text{-}mean)}$
Residual type 2	$TCO_{high\text{-}f} - TCO_{low\text{-}f\ (2nd\ order\ function)}$

Table 3: Summary of sensitivity tests for Pandora relative temperature dependence factors.

Test #	Pandora	Brewer	Data period
1	#103	Combined Brewer (#8, #14, #15, #145, #187, #191)	Oct. 2013 – Dec. 2015
2	#104	Combined Brewer (#8, #14, #15, #187, #191)	Jan. 2014 – Apr. 2014
3	#103	Combined Brewer (#8, #14, #15, #187, #191)	Jan. 2014 – Apr. 2014
4	#103	Combined Brewer (#8, #14, #15, #187, #191)	Oct. 2013 – Aug. 2014
5	#103	Combined Brewer (#8, #14, #145, #191)	Jan. 2015 – Dec. 2015
6	#103	Brewer #191	Oct. 2013 – Dec. 2015
7	#103	Brewer #191	Oct. 2013 – Aug. 2014
8	#103	Brewer #191	Jan. 2015 – Dec. 2015
9	#103	Brewer #8	Oct. 2013 – May. 2015
10	#103	Brewer #14	Nov. 2013 – Dec. 2015
11	#103	Brewer #15	Nov. 2013 – Jul. 2015
12	#103	Brewer #145	Jan. 2015 – Dec. 2015
13	#103	Brewer #187	Oct. 2013 – Apr. 2014
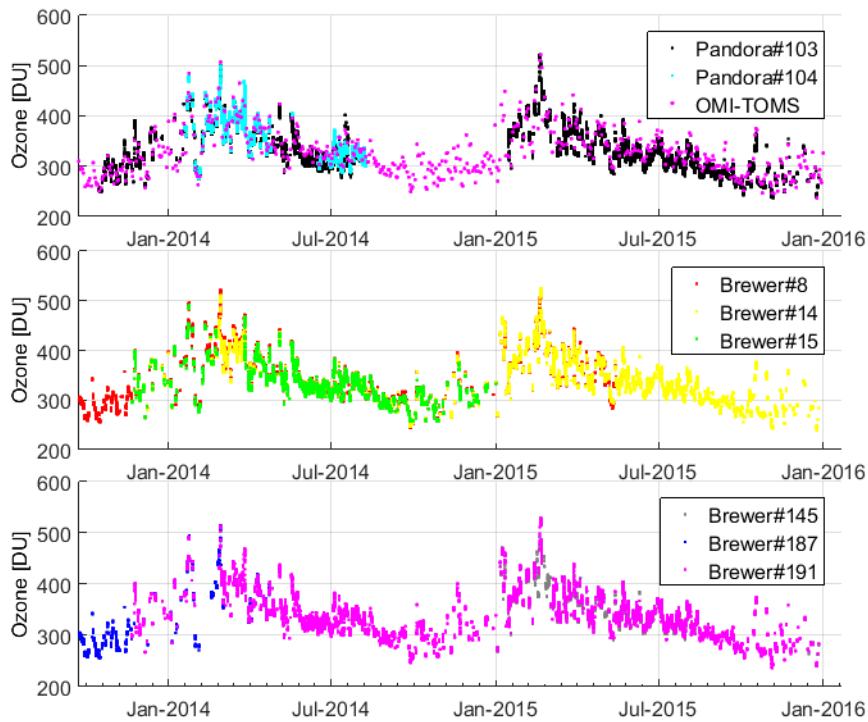
Figure 1: Ozone total column data from Pandoras, Brewers, and OMI.
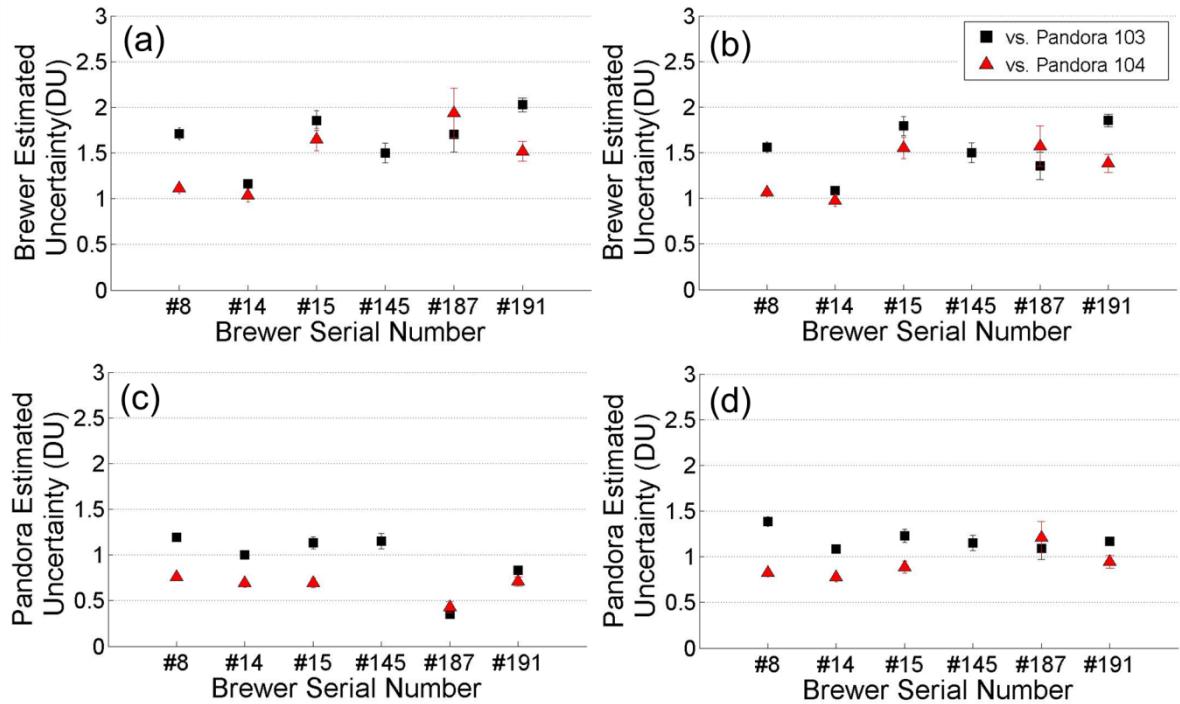
Figure 2: Estimated random uncertainties: for the Brewer instruments using (a) residual ozone type 1, and (b) residual ozone type 2; for the Pandora instruments using (c) residual ozone type 1, and (d) residual ozone type 2. The black squares indicate data from Pandora #103 and the red triangles indicate data from Pandora #104. The error bars show the 95 % confidence bounds.
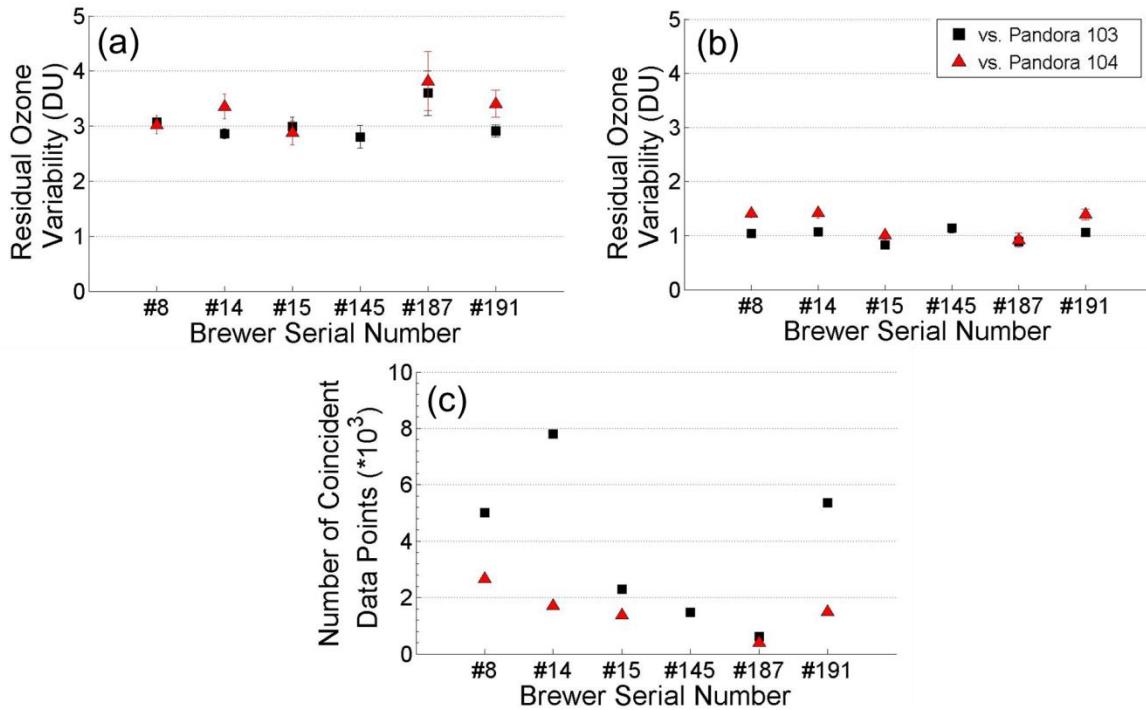
Figure 3 Estimated residual ozone variability (σ_X) using (a) residual ozone type 1, and (b) residual ozone type 2. (c) Number of coincident measurements used in the statistical uncertainty estimation. The black squares indicate data from Pandora #103 and the red triangles indicate data from Pandora #104. The error bars show the 95 % confidence bounds.
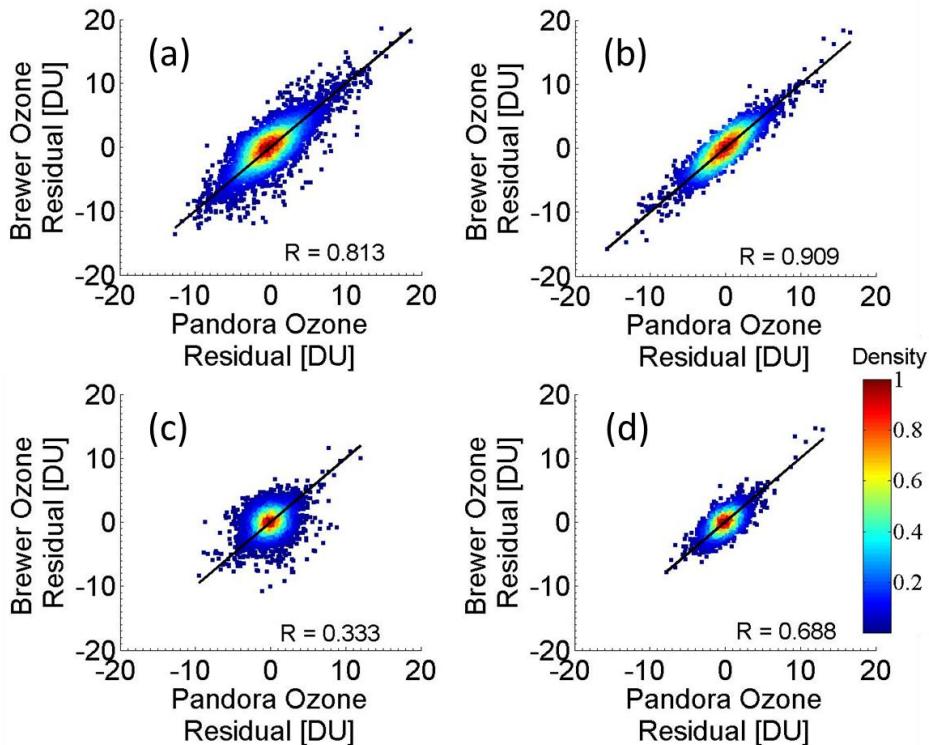
Figure 4: Scatter plots for residual ozone type 1 and 2, colour coded by the normalized density of the points. (a) Brewer #8 vs. Pandora #103 (residual type 1), (b) Brewer #8 vs. Pandora #104 (residual type 1), (c) Brewer #8 vs. Pandora #103 (residual type 2), (d) Brewer #8 vs. Pandora #104 (residual type 2). The black line is the 1-to-1 line.
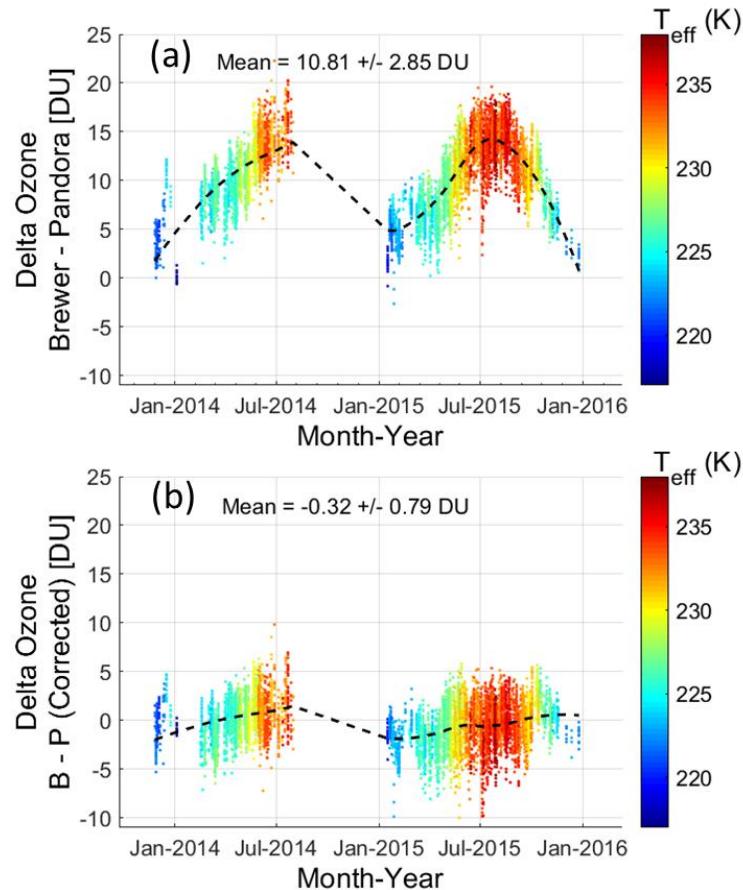
Figure 5: Time series of Brewer #14 – Pandora #103 TCO difference colour coded by ozone effective temperature (see Eq. (1)): (a) before applying the temperature dependence correction, (b) after applying the correction. The dashed lines are Lowess(0.5) fits.
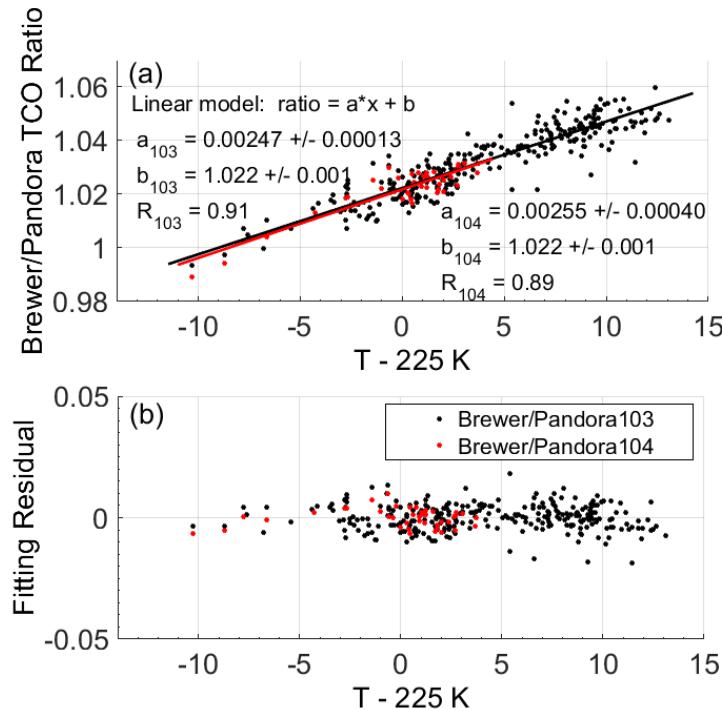
Figure 6: Linear regression of Brewer/Pandora TCO ratio as a function of effective temperature minus 225 K. (a) Linear regression results; (b) residual plot of the linear regression. The black dots indicate data from Brewer/Pandora #103 and the red dots indicate data from Brewer/Pandora #104.
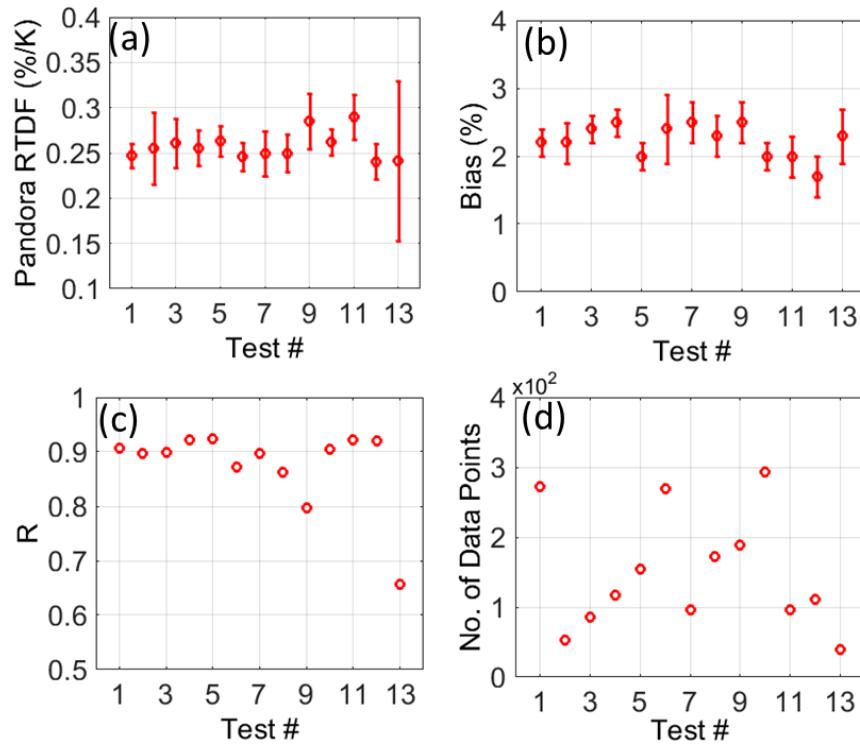
Figure 7: Pandora relative temperature dependence factors derived from 13 sensitivity tests (shown in Table 3). (a) RTDFs, (b) multiplicative biases, (c) correlation coefficients (R), and (d) number of data points in sensitivity tests. The error bars show the 95 % confidence bounds.
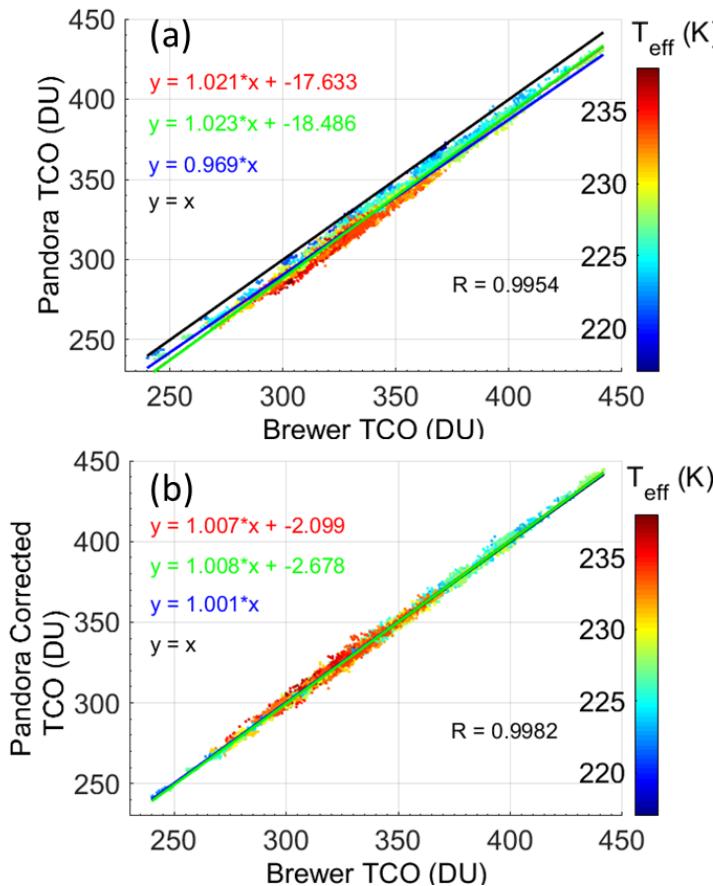
Figure 8: Scatter plots of Pandora #103 vs. Brewer #14 TCO, colour coded by ozone effective temperature: (a) before applying the correction, (b) after applying the correction. The red line is a simple linear fit, the green line is the linear fit weighted by the calculated standard uncertainty from Pandora and Brewer TCO data, the blue line is the linear fit with intercept set to zero, and the black line is the 1-to-1 line.
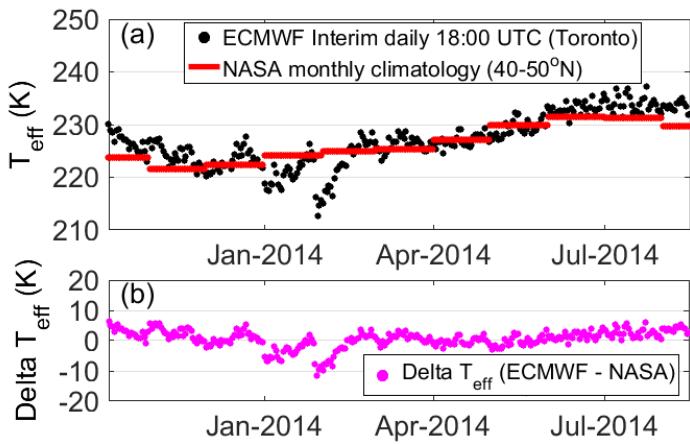
Figure 9: Effective ozone temperature: (a) T*eff* calculated using ECMWF Interim data (18:00 UTC over Toronto) and NASA climatology data (monthly mean for 40-50º N), (b) the difference between these two.
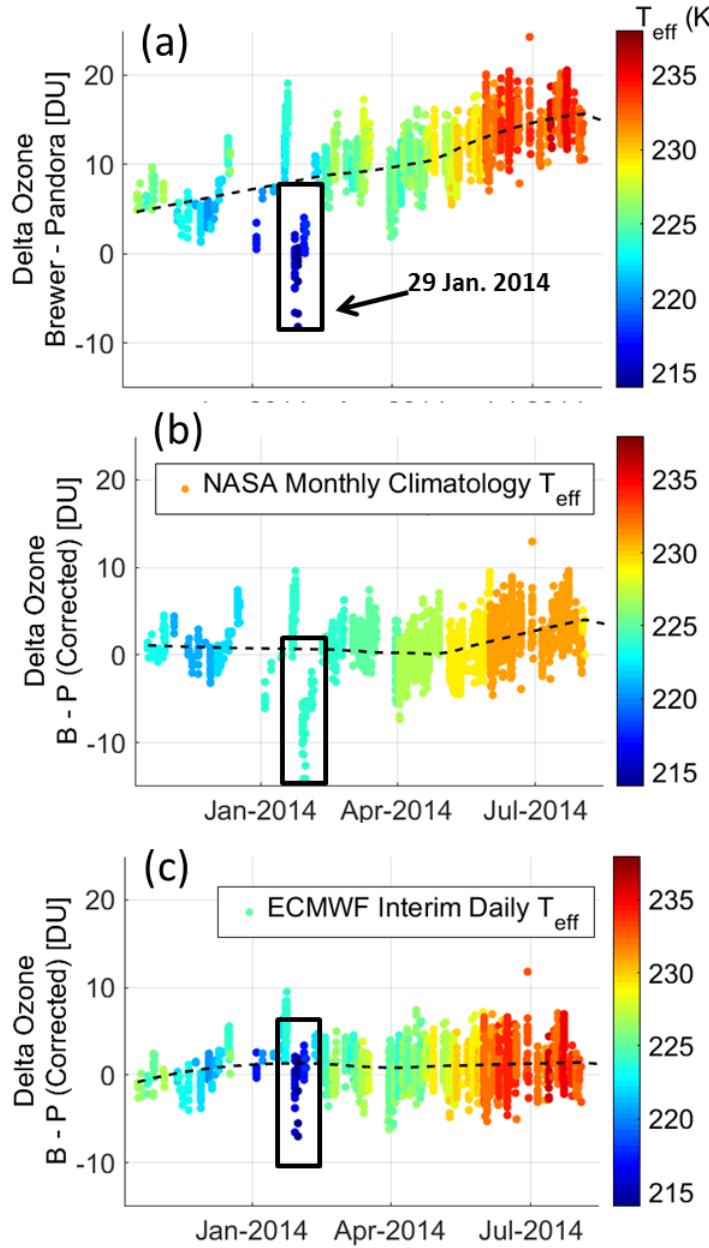
Figure 10: Time series of combined Brewer – Pandora #103 TCO difference colour coded by ozone effective temperature: (a) before applying the temperature dependence correction, (b) after applying the correction using NASA monthly climatology T*eff*, and (c) after applying the correction using ECMWF Interim daily T*eff*. The sudden cooling event on 29-30 January 2014 is marked by black box. The dashed lines are Lowess(0.5) fits.
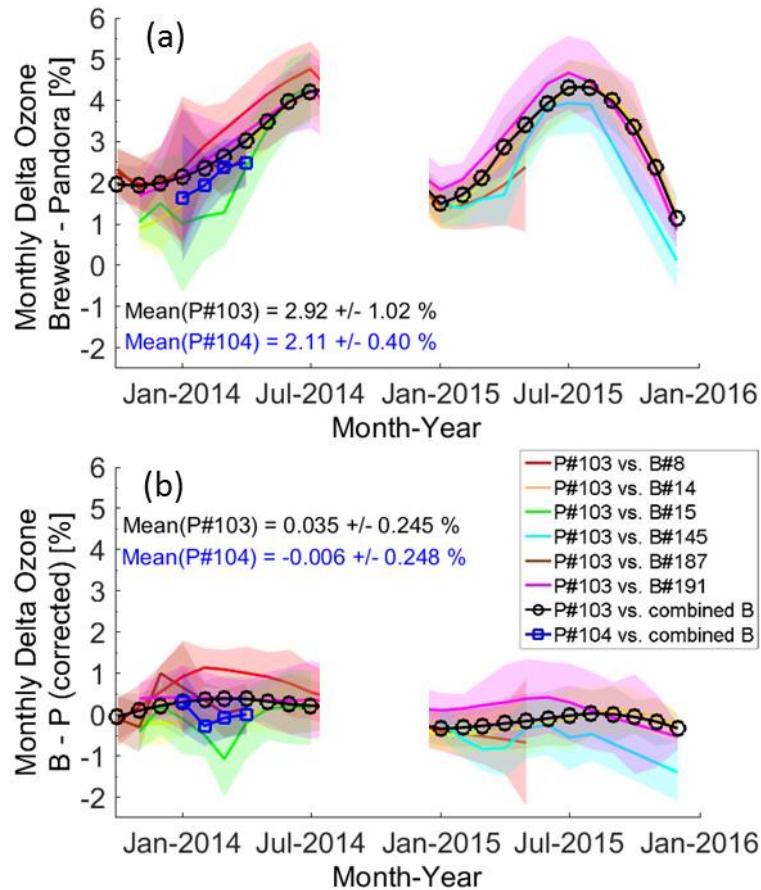
Figure 11: Monthly mean time series of the (Brewer – Pandora)/Brewer % TCO difference: (a) before applying the Pandora temperature dependence correction, and (b) after applying the correction. The shaded regions represent 1σ uncertainty.
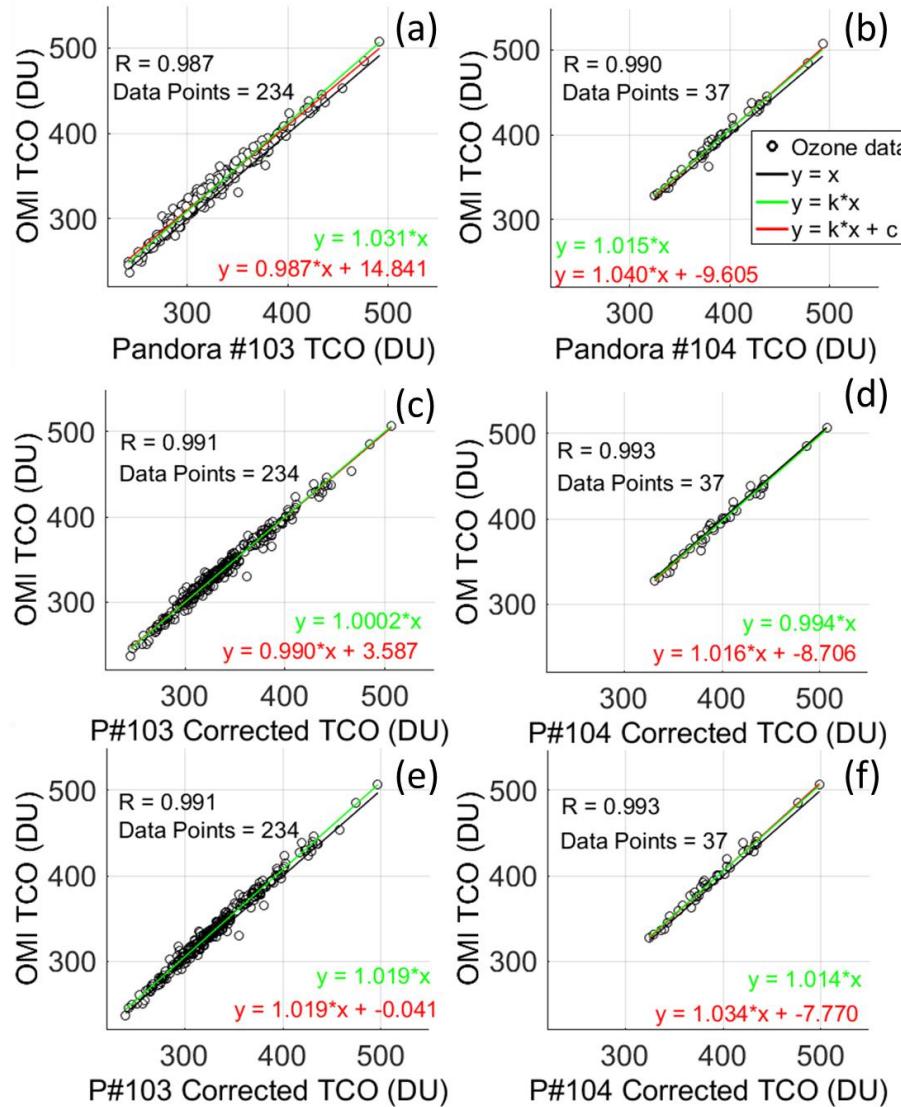
Figure 12: Scatter plots of OMI TCO vs. Pandora TCO for (a) Pandora #103 without TCO correction, (b) Pandora #104 without TCO correction, (c) Pandora #103 with correction using Eq. (11), (d) Pandora #104 with correction using Eq. (12), (e) Pandora #103 with correction using Eq. (13), (f) Pandora #104 with correction using Eq. (13).
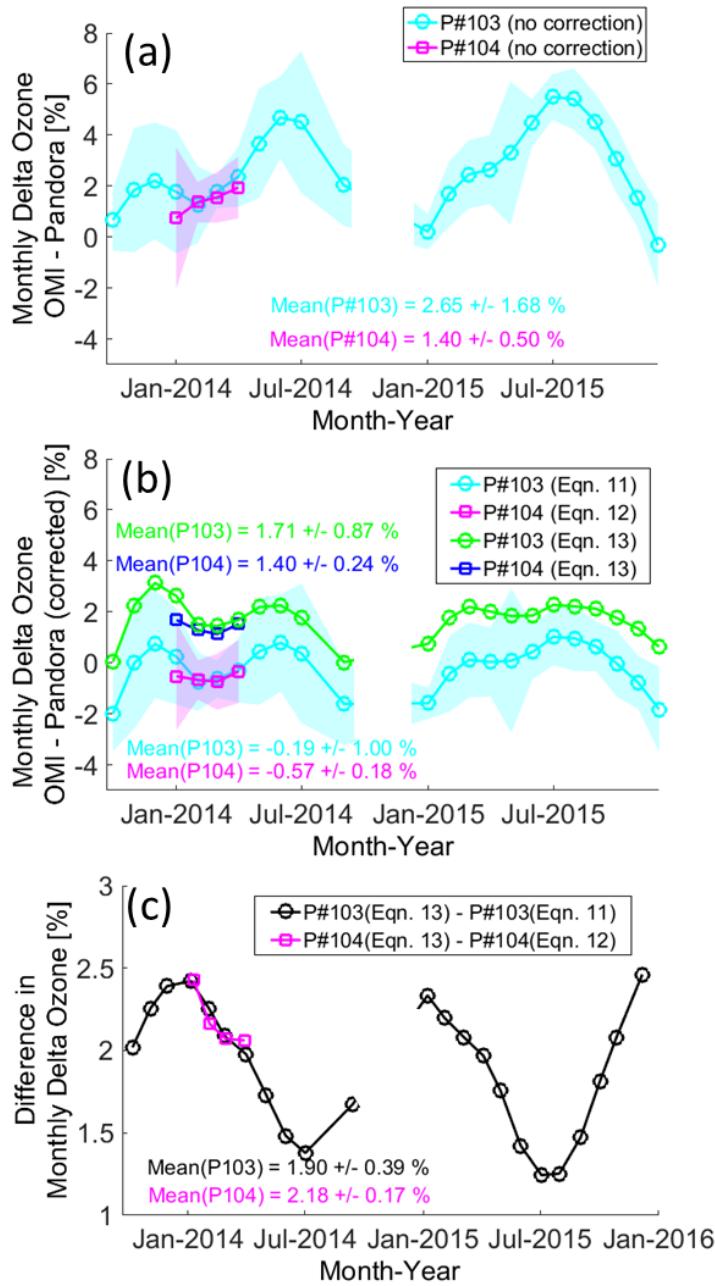
Figure 13: Monthly mean time series of the (OMI – Pandora)/OMI % TCO difference: (a) before applying the correction, (b) after applying the correction using Eqs. (11-13), and (c) the difference between the corrections. The shaded regions represent the 1σ uncertainty.
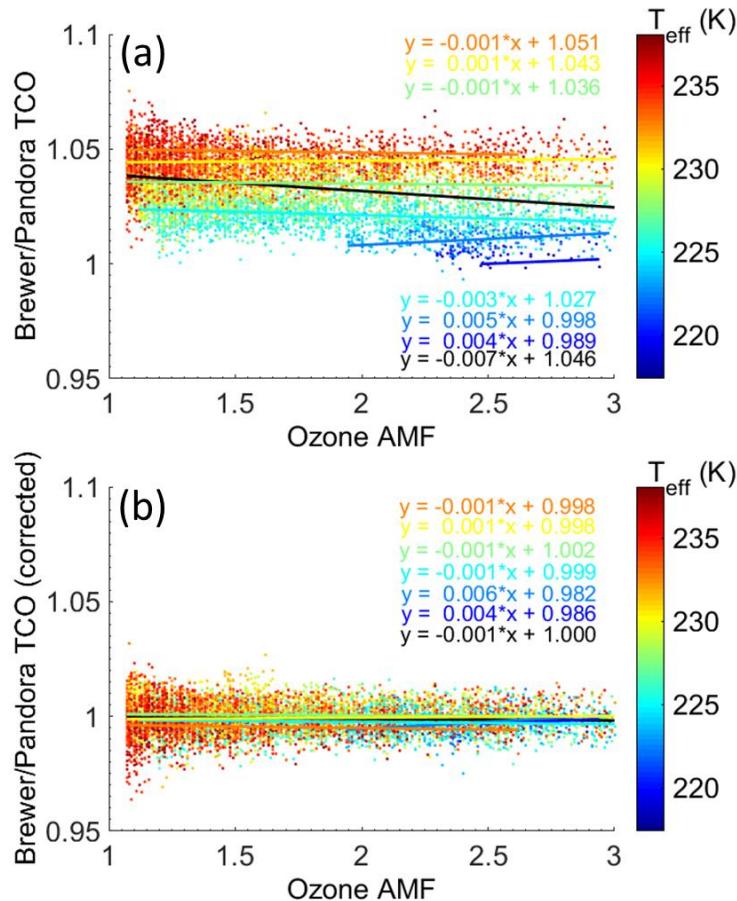
Figure 14: Brewer #14/Pandora #103 TCO ratio vs. ozone airmass factor: (a) before and (b) after applying the Pandora temperature dependence correction. The points are grouped by effective temperature (from 215 to 240 K, in 5 K bins), and the linear fits for each group are colour coded. The black line and linear fit is for the whole dataset.
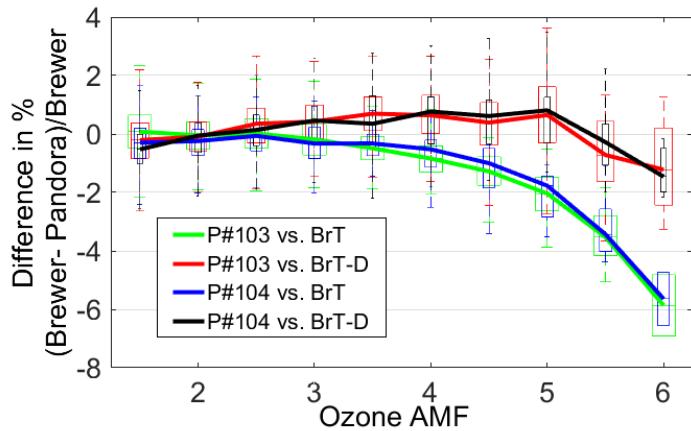
Figure 15: Percentage difference between Pandoras (#103 and #104) and Brewers (grouped as BrT and BrT-D) as a function of ozone airmass factor. On each box, the central mark is the median, the edges of the box are the 25th and 75th percentiles, and the whiskers extend to the most extreme data points not considered outliers.