May 16, 2018

Final response letter to Editor and all Referees of the manuscript "Uncertainty analysis of total ozone derived from direct solar irradiance spectra in the presence of unknown spectral deviations"

Authors: Anna Vaskuri, Petri Kärhä, Luca Egli, Julian Gröbner, and Erkki Ikonen

Article reference: amt-2017-403

Authors: We thank all the Referees once again for their comments. We have taken the comments into account, and after the changes, the manuscript became much better. We believe that our revised version of the manuscript meets the evaluation criteria of the Atmospheric Measurement Techniques journal. We have already given our responses to all referee reports individually that are also attached to this document, but due to the major revision of the manuscript, we list below the main improvements made in the revised manuscript. The revised manuscript with changes highlighted as red and blue colours has been attached after the list of major changes.

This final response letter contains following documents:

- List of the major changes
- Revised manuscript with changes highlighted as red and blue colours
- Response letter to Referee 1
- Two response letters to Referee 2
- Response letter to Referee 3
- Response letter to Referee 4

List of the major changes:

- We further revised the title of our manuscript as it became too long after modifications. The new title is: "Uncertainty analysis of total ozone derived from direct solar irradiance spectra in the presence of unknown spectral deviations".

- We added more background information to the Introduction.

- We revised the title of Section 2 as "ATMOZ field measurement campaign and instrument description" and included more information about the structures and settings of the spectroradiometers used.

- In Section 3, we added a free fitting parameter "*c*" as a scaling factor in our atmospheric model to compensate for relative full spectral correlations. In addition, we changed the least squares minimization method. Due to these improvements, we had to re-analyse all results. The modelled results are now in significantly better agreement, but the uncertainties did not change much as compared with our original manuscript.

- We moved the uncertainty budgets of the three spectroradiometers to a new Section 4.2 “Uncertainty budgets for spectroradiometers”.

- We included a new appendix “Appendix A: Selecting the least squares fitting method”, where we compare the quality of the TOC results obtained with different weightings in the least squares fitting method. Relative least squares fitting with dynamic range limited to four orders of magnitude is now used with QASUME. With BTS and AVODOR, we use a combination of the absolute and relative least squares fitting methods to analyse the results. Using absolute least squares fitting method reduces weight of low irradiance levels that suffer from stray light, thus reducing the systematic inverted U-shape of diurnal TOC results noted with array spectroradiometers.

- We included a new appendix “Appendix B: TOC uncertainties obtained using Chebyshev polynomials”, where we perform the MC uncertainty analysis by using Chebyshev polynomials as base functions and compare the results to those obtained with sinusoidal base functions. The uncertainties do not change significantly when we replace sinusoidal base functions with Chebyshev polynomials.

- In addition, numerous smaller changes have been made in the manuscript according to the referee comments. The response letters we have written and submitted earlier during the discussion phase, listing detailed changes, are included at the end of this response letter after the revised manuscript.

Uncertainty analysis~~Monte Carlo method for determining uncertainty~~ of total ozone derived from direct solar irradiance spectra in the presence of unknown spectral deviations~~: Application to Izaña results~~

Anna Vaskuri[1], Petri Kärhä[1], Luca Egli[2], Julian Gröbner[2], and Erkki Ikonen[1, 3]

[1]Metrology Research Institute, Aalto University, PO Box 15500, 00076 Aalto, Finland
[2]Physikalisch-Meteorologisches Observatorium Davos, World Radiation Center, Dorfstrasse 33, 7260 Davos Dorf, Switzerland
[3]MIKES Metrology, VTT Technical Research Centre of Finland Ltd, PO Box 1000, 02044 VTT, Finland

Correspondence to: Anna Vaskuri (anna.vaskuri@aalto.fi), Petri Kärhä (petri.karha@aalto.fi)

Abstract. We demonstrate a Monte Carlo model to estimate~~calculate~~ the uncertainties of total ozone column~~,~~ (TOC), derived from ground-based direct~~ional~~ solar spectral irradiance measurements. The model estimates the effects of possible systematic spectral deviations in the solar irradiance spectra on the uncertainties in TOC retrieved.~~The model takes into account effects that correlations in the spectral irradiance data may have on the results.~~ The model is tested with spectral data measured with three different spectroradiometers at an intercomparison campaign of the research project "Traceability for atmospheric total column ozone" at Izaña, Tenerife on 17 September 2016. The TOC values derived at local noon have expanded uncertainties of 1.3% (3.6 DU) for a high-end scanning spectroradiometer, 1.5~~3~~% (4.4 DU) for a high-end array spectroradiometer, and 4.7~~3.3~~% (13.3 DU) for a roughly adopted instrument based on commercially available components and an array spectroradiometer when correlations are taken into account. Neglecting the effects of systematic spectral deviations, the uncertainties reduce by a factor of 3. The TOC results of all devices have good agreement with each other, within the uncertainties, and with the reference values of the order of 282 DU during the analysed day, measured with Brewer spectrophotometer #183.~~ The level of TOC measured with reference Brewer spectrophotometer #183 is of the order of 282 DU during the analysed day and in agreement with the results of the two former instruments.~~

1 Introduction

Atmospheric ozone has been defined as an essential climate variable in the global climate observing system (GCOS-200 (2016)) of the World Meteorological Organization (WMO). Its long-term monitoring is necessary to document the expected recovery of the ozone layer due to the implementation of the Montreal protocol (UNTC (1987)) and its amendments on the protection of the ozone layer. Atmospheric ozone, first discovered by Fabry and Buisson (1913), protects the humans, the biosphere, and infrastructures from adverse effects of ultraviolet (UV) radiation by shielding the Earth surface from excessive radiation levels (McElroy and Fogal (2008)). Since the 1970's, it is known that human-produced chlorofluorocarbons (CFCs)

destruct atmospheric ozone (Molina and Rowland (1974)) and have led to recurring massive losses of total ozone in the Antarctic in the form of the ozone hole (Farman *et al.* (1985); Solomon *et al.* (1986)). An unprecedented ozone depletion has also been recently observed in the Arctic (Manney *et al.* (2011)), while in the middle-latitudes, moderate ozone depletion has been observed (Solomon (1999)). The Montreal protocol and its amendments have been successful in reducing the emission of ozone-depleting substances (Velders *et al.* (2007)). Nevertheless, recent studies give conflicting results with respect to the observation of a recovery of the ozone layer, and model projections have shown the recovery to occur not before the middle of the 21st century (Ball *et al.* (2018); Weber *et al.* (2018)). Therefore, careful monitoring of the thickness of the ozone layer with uncertainties of 1% or less is crucial in verifying the successful implementation of the Montreal Protocol and the eventual recovery of the ozone layer to pre-1970's levels.

"Traceability for atmospheric total column ozone" (ATMOZ) was a three-year project funded partly by the European Metrology Research Programme (EMRP) and the European Union (ATMOZ project (2014 – 2017)). The goal of this project was to produce traceable measurements of total ozone column (TOC) with uncertainties down to 1%, by a systematic investigation of the radiometric and spectroscopic aspects of the methodologies used in retrieval. Another objective of the project was to provide a comprehensive treatment of uncertainties of all parameters affecting the TOC retrievals using spectrophotometers. This paper presents outcome of the work on studying the uncertainty of TOC obtained from spectral direct solar irradiance measurements, taking unknown spectral errors explicitly into account.

~~Total ozone column (~~TOC~~)~~ can be determined from spectral measurements of direct solar ~~ultraviolet (~~UV~~)~~ irradiance (Huber *et al.* (1995)). We have developed a Monte Carlo (MC) based model to estimate the uncertainties of the derived TOC values. One frequently overlooked problem with uncertainty evaluation is that the spectral data may hide systematic wavelength dependent errors due to unknown correlations (Kärhä *et al.* (2017b, 2018); Gardiner *et al.* (1993)). ~~The reason often is that the correlations are not known.~~ Omitting possible correlations may lead into underestimated uncertainties for derived quantities, since spectrally varying systematic errors typically produce larger deviations than uncorrelated noise-like variations that traditional uncertainty estimations predict. Complete uncertainty budgets for quantities measured are necessary to understand~~Accounting for increased uncertainties arising from the correlations improves reliability of observing and predicting~~ long term environmental trends, such as changes in the stratospheric ozone concentration (e.g. Molina and Rowland (1974)) and solar UV radiation (e.g. Kerr and McElroy (1993); McKenzie *et al.* (2007)).

Physically, spectral correlations may originate, e.g., from lamps or other light sources used in calibrations. If their temperatures change e.g. due to ageing or current setting, a spectral change in the form of Planck's radiation law is introduced. Non-linearity in the responsivity of a detector causes systematic differences between high and low measured values. The introduced spectrally systematic but unknown changes in irradiance may change the derived TOC values significantly, exceeding the uncertainties calculated assuming that the uncertainty in irradiance behaves like noise. The presence of correlations in measurements can be seen in many ways. For example, problems have occurred when new ozone absorption cross-sections have been taken into use (Redondas *et al.* (2014); Fragkos *et al.* (2015)). Derived ozone values may change significantly because different systematic errors are included in the different cross-sections. Also, TOC estimated from a

measured spectrum often depends on the wavelength region chosen, although the measurement region should not affect the result much.

In this paper, we introduce a new method for dealing with possible correlations in spectral irradiance data and analyse uncertainties in ozone retrievals for three different spectroradiometers used in a recent ATMOZ intercomparison campaign at Izaña, Tenerife, to demonstrate how it can be used in practice. One of the instruments is ~~the~~ QASUME ~~spectroradiometer~~ (Gröbner *et al.* (2005)) that ~~serves as a high-quality reference instrument of spectral irradiance~~is the World reference UV spectroradiometer at the World Radiation Center (PMOD/WRC). The second one is an array-based high-quality spectroradiometer BTS2048-UV-S-WP (BTS) from Gigahertz-Optik ~~AVODOR, a simple array based spectrograph from Avantes, and the third one is BTS2048-30-UV-S-WP (BTS) from Gigahertz-Optik, an array-based high-quality spectrograph that has been optically corrected for stray light~~(Zuber *et al.* (2017a, b)), operated by Physikalisch-Technische Bundesanstalt (PTB). The third one is an array-based spectroradiometer AvaSpec-ULS2048LTEC (AVODOR) from Avantes, operated by PMOD/WRC. The field of view of the spectroradiometers has been limited so that they measure direct~~ional~~ spectral irradiance of the Sun ~~only~~, excluding most of the indirect radiation from the remainder of the~~blue~~ sky.

2 ATMOZ~~The Izaña~~ field measurement campaign and instrument description

~~Traceability for atmospheric total column ozone (ATMOZ) was a three-year research project funded by the EMRP program of the European Union in 2014 – 2017.~~ The ATMOZ project arranged a field measurement campaign (ATMOZ campaign (2016)) that took place ~~between~~ 12 – 30~~25~~ September 2016 at the Izaña Atmospheric Observatory, a high mountain Global Atmospheric Watch (GAW) station located at an altitude of 2.36 km above the sea level in Tenerife, Canary Islands, Spain (28.3090° N, 16.4990° W). The measurement campaign was organised by the Spanish Meteorological Agency (AEMET) and the ~~World Radiation Center (~~PMOD/WRC~~)~~ for the intercomparison of TOC measured with different participating instruments, including Dobson and Brewer spectrophotometers, and various spectroradiometers.

The focus of this paper is to study uncertainties of the TOC values derived from direct solar UV irradiance spectra. Total ozone column TOC is the vertical ozone profile $\rho_{O_3}(z)$ integrated over altitudes as

$$\mathrm{TOC} = \int_{z_0}^{z_1} \rho_{O_3}(z)\,\mathrm{d}z \tag{1}$$

from the station altitude z_0 to the top-of-the-atmosphere altitude z_1. We study the data measured during the day of 17 September 2016 with three different spectroradiometers using the ozone retrieval algorithm introduced in Section 3. ~~The measurements took place on the mountain Teide at the altitude of 2.36 km above the sea level (28.3090° N, 16.4990° W).~~ Station pressure ~~of 772.8 hPa~~ was monitored during the campaign and determined to be 772.8 hPa with a standard uncertainty of 1.3 hPa. The ozone and temperature profiles were measured with a sonde ~~prior to measurements as~~ during the campaign and examples of them are shown in Fig. 1.

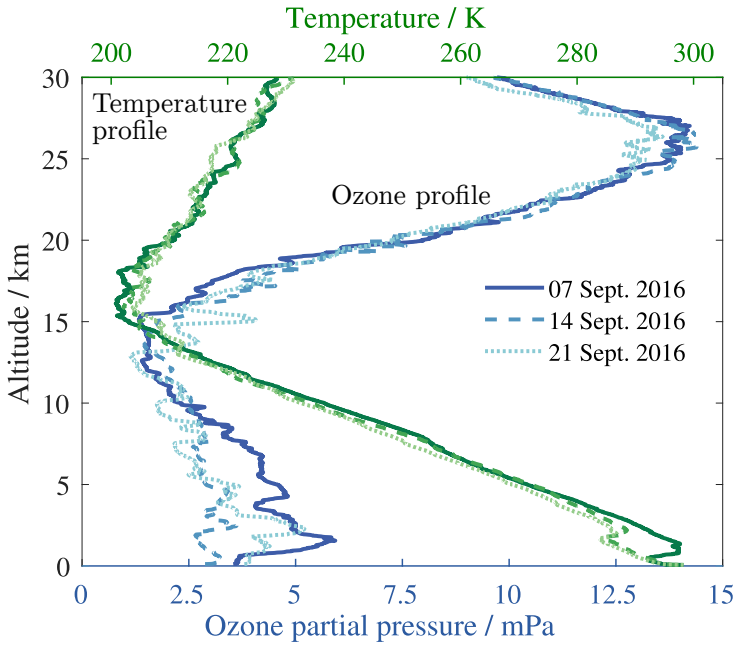


Figure 1. Temperature and ozone profiles as a function of the altitude, measured with a sonde during the ATMOZ field measurement campaign.

OIn our ozone retrieval method uses one atmospheric layer, vertical profiles were not implemented in the model to reduce computational complexity, which may slightly shift the absolute values, but should not have an effect on the uncertainties obtained. With thea one layer model, the ozone absorption cross-section is a function of the effective temperature and the relative air mass is a function of the effective altitude of the ozone layer. Using the vertical ozone profile $\rho_{O_3}(z)$, the effective altitude $h_{\text{eff}} = 26$ km $\pm$ 0.5 km of the ozone layer was estimated by integration over altitudes

$$h_{\text{eff}} = \frac{\int_{z_0}^{z_1} z\, \rho_{O_3}(z)\, \mathrm{d}z}{\int_{z_0}^{z_1} \rho_{O_3}(z)\, \mathrm{d}z}, \tag{2}$$

from the station altitude z_0 to the top-of-the-atmosphere altitude z_1. The corresponding effective temperature $T_{\text{eff}} = 228$ K $\pm$ 1 K was estimated (Komhyr *et al.* (1993); Fragkos *et al.* (2015)) as

$$T_{\text{eff}} = \frac{\int_{z_0}^{z_1} T(z)\, \rho_{O_3}(z)\, \mathrm{d}z}{\int_{z_0}^{z_1} \rho_{O_3}(z)\, \mathrm{d}z}. \tag{3}$$

The uncertainties stated for $h_{\text{eff}} = 26\ \text{km} \pm 0.5\ \text{km}$ and $T_{\text{eff}} = 228\ \text{K} \pm 1\ \text{K}$ are standard deviations. The values and uncertainties are estimated from the measured vertical profiles (Fig. 1). The profiles are extended to 80 km using climatology.

The data sets measured by three different spectroradiometers were studied in this work. These spectroradiometers use different techniques to measure the spectral distribution of radiation. Monochromator-based spectroradiometers, such as QASUME, measure one nearly monochromatic wavelength band at a time, and thus measuring the full spectrum is relatively

slow. On the other hand, they usually have significantly better stray light properties than array-based spectroradiometers, such as BTS and AVODOR that image the full spectrum at once by dispersing the incoming radiation towards a photodiode array.

~~One of the instruments was the QASUME double-monochromator spectroradiometer (Gröbner et al. (2005)) of PMOD/WRC.~~QASUME spectroradiometer collects and guides the incoming radiation with input optics and a quartz fiber bundle to the entrance slit of a Bentham DM150 double monochromator (Gröbner *et al.* (2005)). One wavelength at a time is selected by rotating the two gratings of the double-monochmomator. Then, the monochromatic signal is measured with a photomultiplier tube. QASUME is usually operated in global spectral irradiance mode (Gröbner *et al.* (2005); Hülsen *et al.* (2016)), but during the campaign it was equipped with a collimator tube with a full opening angle of 2.5° allowing the measurement of direct solar spectral irradiance (Gröbner *et al.* (2017)). The measurement range of QASUME during the campaign was limited to 250 nm – 500 nm with a step interval of 0.25 nm, so that one spectrum was~~could be~~ measured every 15 minutes. To ensure stable outdoor measurements, the double-monochromator of QASUME was mounted inside a temperature-controlled weather-proofed box (Hülsen *et al.* (2016)).

BTS spectroradiometer utilizes a stationary grating and a back-thinned cooled CCD array detector, mounted in a Czerny-Turner configuration~~The second one was AVODOR, a simple array based spectrograph from Avantes, with a step interval of 0.14 nm. The third one was an array-based spectrograph from Gigahertz-Optik that has been optically corrected for stray light using an internal filter wheel~~ (Zuber *et al.* (2017a, b))~~, operated by PTB~~. To measure direct solar spectrum, BTS was equipped with a collimator tube with a full opening angle of 2.8° designed by PTB, and it uses an internal filter wheel system with eight filter positions together with a specific measurement routine to reduce stray light. BTS was mounted on a solar tracker, EKO STR-32G by EKO Instruments Co., Ltd., with pointing accuracy better than 0.01°. A weather-proof housing with temperature control allows BTS operation at the ambient temperatures from –25 °C to 50 °C. During the ATMOZ campaign, the housing temperature of BTS was measured to be stable within 0.1°C (Zuber *et al.* (2017a, b)). The measurement range of BTS was 200 nm – 430 nm with a step size of 0.2 nm during the campaign. One spectrum was measured every 45 seconds. ~~The measurement range of BTS was between 200 nm and 430 nm with a step interval of 0.2 nm during the campaign.~~

AVODOR spectroradiometer has a stationary grating and a back-thinned cooled CCD array detector in a Czerny-Turner configuration. AVODOR measures the spectrum from 200 nm to 540 nm with a step size of 0.14 nm in the UV region. During the ATMOZ campaign, the field of view of AVODOR was limited to 1.5° by a commercial collimator tube used, J1004-SMA by CMS Ing.Dr.Schreder GmbH. The spectral range of AVODOR was limited between 295 nm and 345 nm by a combination of two solar blind filters to reduce stray light from the visible and infrared parts of the solar spectrum. The solar blind filters were mounted between the collimator tube and the fiber entrance of the spectroradiometer. One spectrum was measured every 30 seconds.

~~Uncertainty components of the spectral measurements are listed in Table 1 for QASUME, in Table 2 for AVODOR, and in Table 3 for BTS spectroradiometers. The tables also give division of the uncertainty components to different correlation types as described in Section 4.~~

The slit functions of the three spectroradiometers shown in Fig. 2 were measured with lasers before the measurement campaign. They are needed when fitting the modelled spectra at the Earth surface ~~the ozone retrieval~~ to the measured spectra.

In addition, it is of importance to notice the different wavelength steps of the data, 0.25 nm for QASUME, 0.2 nm for BTS, and 0.14 nm for AVODOR. The wavelength steps of the spectral data affect the magnitude of the uncertainties in TOC created by spectrally random components. In our full spectrum TOC retrieval, t~~T~~he number of data points n~~N~~ which is smaller with a larger wavelength step interval, affects uncertainties with a factor of $1/\sqrt{n}$~~$\sqrt{N}$~~ as demonstrated in (Kärhä *et al.* (2017b); Poikonen *et al.* (2009)).

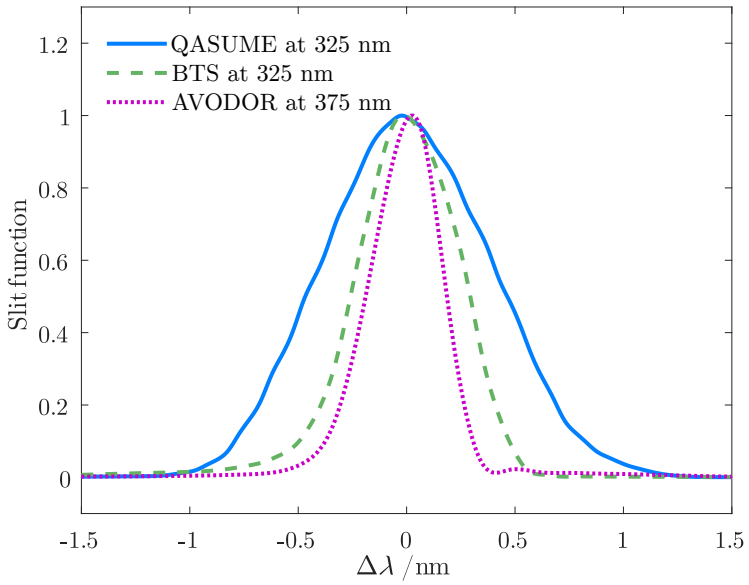


Figure 2. Slit functions, measured with narrow band lasers, for the spectroradiometers used in the Izaña campaign. The laser wavelengths are stated in legends.

~~This approach is consistent with the ozone measurement with~~Brewer MkIII spectrophotometers~~, which are~~ used as reference devices for ozone measurements established by the International Meteorological Organization, the Commission for Instruments and Methods of Observation (CIMO) (Redondas *et al.* (2016)), also measure the spectral irradiance at UV region, but using four narrow wavelength bands at 310.1 nm, 313.5 nm, 316.8 nm, and 320.0 nm (Kipp & Zonen (2015)). The Brewer MkIII instruments solve absolute TOC values ~~through rearrangement of Eq. (3) and~~ by comparing the logarithms of ratios of count rates between four wavelength channels, i.e. using the double ratio technique. Determining the TOC using the double ratio method is invariant for such spectral deviations which have the same relative magnitude at all wavelengths, i.e spectrally constant deviations.

Our full spectrum retrieval method performs averaging in the wavelength domain, whereas Brewer spectrophotometer does it in the time domain. Brewer can measure up to tens of seconds to get millions of photons, so that the photon noise reduces to a level of 0.1%. At this low noise level, it is not critical that only four wavelengths are used. As Brewer instruments are well-known and widely used, we also compare TOC results obtained using our full spectrum retrieval method and the spectroradiometers to those measured by Brewer #183 during the same day.

3 Atmospheric model

In this study, we use an atmospheric ozone retrieval algorithm in many aspects similar to the article by Huber *et al.* (1995). The relationship between the ~~terrestrial~~ spectral irradiance $E_{\mathrm{s}}(\lambda)$ at the Earth surface and the extra-terrestrial solar spectrum $E_{\mathrm{ext}}(\lambda)$ is based on Beer–Lambert–Bouguer absorption law (Beer (1852); Lambert (1760); Bouguer (1729))~~defined~~ as

$$E_{\mathrm{s}}(\lambda) = c \cdot E_{\mathrm{ext}}(\lambda) \cdot \exp\left[-\alpha_{\mathrm{O_3}}(\lambda, T_{\mathrm{eff}}) \cdot \mathrm{TOC} \cdot m_{\mathrm{TOC}} - \tau_{\mathrm{R}}(\lambda, P_0, z_0, \phi) \cdot m_{\mathrm{R}} - \tau_{\mathrm{AOD}}(\lambda) \cdot m_{\mathrm{AOD}}\right], \tag{4}$$

where $\alpha_{\mathrm{O_3}}(\lambda, T_{\mathrm{eff}})$ is the ozone absorption cross-section at the effective ozone temperature T_{eff}, $\tau_{\mathrm{R}}(\lambda, P_0, z_0, \phi)$ is the Rayleigh scattering optical depth that depends on the station pressure P_0, the station altitude z_0, and the geographic latitude ϕ. The QASUME-FTS data set by Gröbner *et al.* (2017) was used as the extra-terrestrial spectrum $E_{\mathrm{ext}}(\lambda)$. Parameter c is a scaling factor set as a free parameter to compensate for spectrally constant deviations.

The relative air mass of the ozone layer with the Earth curvature taken into account can be expressed as

$$m_{\mathrm{TOC}} = \frac{1}{\cos\left[\arcsin\left(\frac{R}{R+h_{\mathrm{eff}}} \cdot \sin\theta\right)\right]}, \tag{5}$$

where θ is the incident solar zenith angle at the observing site~~between vacuum-to-air interface~~ that is a function of the local time, date, and geographic coordinates (Meeus (1998); Reda and Andreas (2008)). Parameter h_{eff} is the altitude of the ozone layer from the ground, and R is the radius of the Earth. As the ozone and other molecules creating ~~the Rayleigh~~ scattering are distributed at different altitudes, we calculate the relative air mass factor m_{R}~~s $m_{\mathrm{R}} \approx m_{\mathrm{AOD}}$~~ for Rayleigh scattering~~and aerosols~~ at the altitude of 5 km (Gröbner and Kerr (2001)) and approximate the relative air mass factor of aerosols so that $m_{\mathrm{AOD}} \approx m_{\mathrm{R}}$ (Gröbner *et al.* (2017)). The temperature dependence of $\alpha_{\mathrm{O_3}}(\lambda, T_{\mathrm{eff}})$ between 203 K and 253 K (Weber *et al.* (2016)) was interpolated by a second degree polynomial at each wavelength (Paur and Bass (1985)) as

$$\alpha_{\mathrm{O_3}}(\lambda, T_{\mathrm{eff}}) = a_1(\lambda) T_{\mathrm{eff}}^2 + a_2(\lambda) T_{\mathrm{eff}} + a_3(\lambda). \tag{6}$$

We take the Rayleigh scattering optical depth into account by the state-of-the-art model by (Bodhaine *et al.* (1999)). The aerosol optical depth (AOD) is approximated from the Ångström AOD model (Ångström (1964)) as

$$\tau_{\mathrm{AOD}}(\lambda) = \beta \cdot \left(\frac{\lambda}{1\ \mu\mathrm{m}}\right)^{-\alpha}, \tag{7}$$

where constant $\alpha \approx 1.4$ is the Ångström coefficient at typical atmospheric conditions and $\beta \geq 0$ is the Ångström turbidity~~extinction~~ coefficient. ~~The wavelength λ is expressed in units of µm.~~

The model~~terrestrial~~ spectrum $E_{\mathrm{s}}(\lambda)$ at the Earth surface, convolved by the slit function of the spectroradiometer~~,~~ as shown in Fig. 2, is fitted with parameters TOC,~~and~~ β, and c to the measured ground-based spectrum $E(\lambda)$. The absolute TOC level obtained ~~varies~~ depend~~sing~~ on the ~~selected least squares~~ fitting method used. We used a ~~relative~~ least squares fitting method (Levenberg (1944); Marquardt (1963)) with trust-region optimisation by Matlab function 'lsqnonlin' as

$$S = \sum_{i=1}^{n} w(\lambda_i) \cdot \left[E_{\mathrm{s}}(\lambda_i) - E(\lambda_i)\right]^2, \tag{8}$$

where S is the sum of the squared residuals to be minimised, and $w(\lambda_i)$ is the weight for each point measured. Index $i = 1, 2, ..., n$ runs over the wavelengths of the spectral measurements.

~~There are various ways of how to deal with noise, and the selection affects the obtained fitting results. In this work, we limit the spectral range for each device by selecting a wavelength region where the signal is above the noise floor in all measurements. Removing noisy results improves the quality of fits especially with the spectrograph type instruments AVODOR and BTS that suffer from higher stray light than QASUME. Generally, measurement results do not improve by reducing data points, and only the data that are below the noise floor should be removed. Other possibilities to deal with noise include weighting the measurement results with the inverse square of uncertainties $u(\lambda)^{-2}$. In addition, a noise floor of $5 \cdot 10^{-3}$ was used for AVODOR and BTS, removing all data below this level, which affect measurements in the morning and in the evening. With QASUME, the noise floor was set to 10^{-5}. These selections can be seen in the modelled spectra (black solid curves) in Fig. 3.~~

Figure 3 presents examples of measured~~ments~~ and modelled spectra~~values~~ for the spectroradiometers used in this work. As can be seen, the signal-to-noise ratios and stray light properties of the devices differ significantly among different spectroradiometers~~ from each other~~. All spectra measured by QASUME are practically noiseless above 10^{-6} $\mathrm{Wm^{-2}nm^{-1}}$, resulting in a dynamic range of approximately four orders of magnitude. The dynamic range for BTS is approximately two orders of magnitude and less than two orders of magnitude for AVODOR.

For QASUME, we use the relative least squares fitting method (RLS) minimisation with $w(\lambda) = E(\lambda)^{-2}$, as QASUME does not suffer from stray light and RLS fitting has been used in the past for monochromator-based spectroradiometers, e.g. by Huber *et al.* (1995). These least squares fitting selections are discussed in more detail in Appendix A.

To minimise the effect of stray light, we use absolute least squares minimisation, also known as ordinary least squares fitting method (OLS), with $w(\lambda) = 1$ for BTS and AVODOR, as this selection is less affected by the lowest irradiance levels where the stray light and noise are dominant. As shown in Appendix A, using OLS introduces an offset, approximately 1% in these measurements, to the retrieved TOC values. We take this into account by analysing the results at noon that are less influenced by stray light also with RLS, and make a correction to the OLS results as

$$\mathrm{TOC_{UTC,OLS,c}} = \frac{\mathrm{TOC_{noon,RLS}}}{\mathrm{TOC_{noon,OLS}}} \cdot \mathrm{TOC_{UTC,OLS}}. \quad (9)$$

After this correction, results of all instruments, QASUME, BTS, and AVODOR are comparable.

The shortest fitting~~modelling~~ wavelength for the~~ monochromator based~~ spectroradiometers ~~(QASUME)~~ in this work was selected to be 300 nm since the typical stray light compensation methods~~rrections~~ are not effective below 300 nm (Nevas *et al.* (2014)). ~~The same limit was used for the stray light corrected BTS array spectrometer. AVODOR array spectrometer does not have any stray light correction, it shows significantly larger noise at lower wavelength, and therefore we set the short wavelength limit to 305 nm.~~ The upper fitting wavelength limit was set to 340 nm with all three spectroradiometers as the ozone absorption is not effective above that wavelength. Due to the relatively large bandwidths of the spectroradiometers (Fig. 2), calculations before the convolution and the convolution itself were carried out over a wider range 295 nm – 345 nm.

To see whether a global optimum is achieved with our atmospheric ozone retrieval method, we varied the initial guess values of TOC from 10 DU to 700 DU, β from 0 to 0.5, and c from 0 to 100. Within the ranges stated, the free parameters always converged to the same final values regardless of the initial guess values.

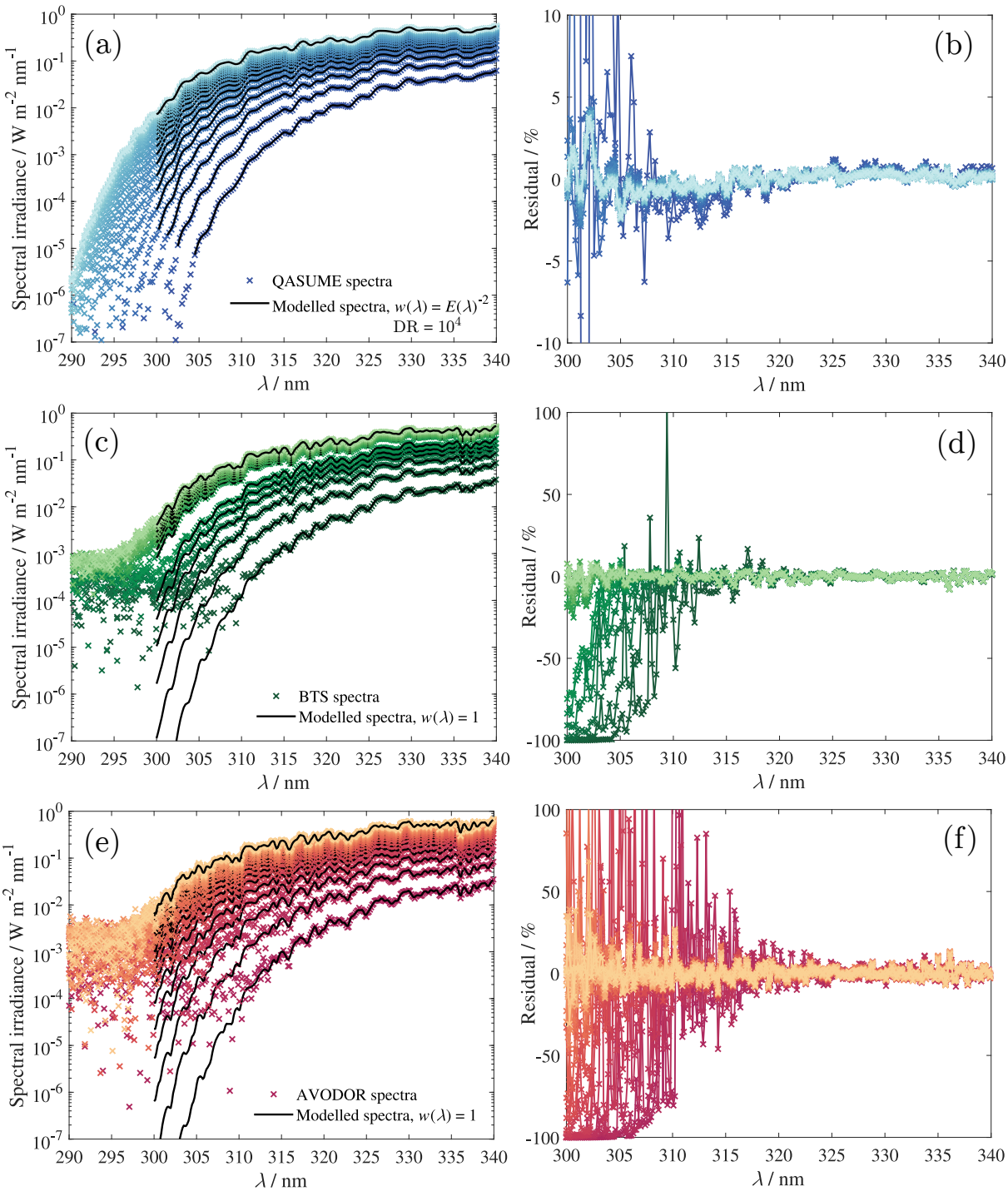


Figure 3. Examples of fitting the atmospheric model to the direct ground-based solar UV spectra between 300 nm and 340 nm for QASUME (a–b) , BTS (c–d~~b~~) and AVODOR (e–f~~c~~). In figures on the left hand side, t~~T~~he coloured symbols indica~~deno~~te measured spectra, and the black solid curves indicate modelled spectra. Figures on the right hand side show the relative spectral residuals of the fits. In (a), the abbreviation DR refers to the dynamic range of QASUME data used in the least squares fitting.

4 Uncertainty estimation

4.1 Uncertainty model

In uncertainty analysis, the combined uncertainties are calculated with the square sum of the standard deviations of the components, i.e. their variances are summed up. If correlations of uncertainties are known, they should be taken into account. This can be carried out with the methods defined in the Guide to the Expression of Uncertainty in Measurement (JCGM 100 (2008)). In this paper, we do this for all uncertainty components, where the mechanism of contributing to the uncertainty of TOC is known. However, with some of the components, we do not know exactly the mechanisms leading into correlations. With such uncertainties, we estimate the effects that possible correlations might have using a newly developed MC model described in (Kärhä *et al.* (2017a, b)).

In our MC model, p~~P~~ossible systematic deviations within uncertainties are reproduced using a cumulative Fourier series

$$\delta(\lambda) = \sum_{i=0}^{N} \gamma_i f_i(\lambda) \tag{10}$$

with sinusoidal base functions, shown in Fig. 4, as

$$f_i(\lambda) = \sqrt{2} \cdot \sin\left[i \left(2\pi \frac{\lambda - \lambda_1}{\lambda_2 - \lambda_1} \right) + \phi_i \right], \tag{11}$$

where index $i = 1, 2, ..., N$ depicts the order of complexity of the deviation (Kärhä *et al.* (2017b));~~,~~ λ_1~~= 300 nm or 305 nm depending on the spectroradiometer~~ and λ_2~~= 340 nm~~ limit the wavelength range of the analysis. For calculations before the convolution due to slit function, this range was set to 295 nm – 345 nm. Otherwise, the results at the ends of the range 300 nm – 340 nm would be distorted due to incomplete convolution. This concerns e.g. uncertainty of the extra-terrestrial spectrum and the ozone absorption cross-section. After the convolution, actual fitting of the modelled spectra to the measured spectra was carried out at 300 nm – 340 nm, and this range was also used in the uncertainty analysis of the components related to the measured spectra. T~~, and t~~he phase ϕ_i of the base function is an equally distributed MC variable between 0 and 2π. In addition, $f_0(\lambda) = 1$ is used to account for constant offset. The weights γ_i for the base functions are selected in an N-dimensional spherical coordinate system (Hicks and Wheeling (1959)) in such a way that the variance of the final deviation~~error~~ function always equals to~~remains~~ unity. In practice, this means that the weights γ_i are generated by scaling the random variables $Y_i \sim \mathcal{N}(0,1)$ as

$$\gamma_i = \frac{Y_i}{\sqrt{Y_0^2 + Y_1^2 + ... + Y_N^2}}. \tag{12}$$

The complete N-dimensional system requires orthogonal base functions, such as full periods of sine functions, to allow an arbitrary shape of deviation function $\delta(\lambda)$ with unity variance. It is possible to use other orthogonal sets of functions, such as Chebyshev polynomials instead of sinusoidal base functions, but that would involve more complicated mathematics. This is discussed in more detail in Appendix B.

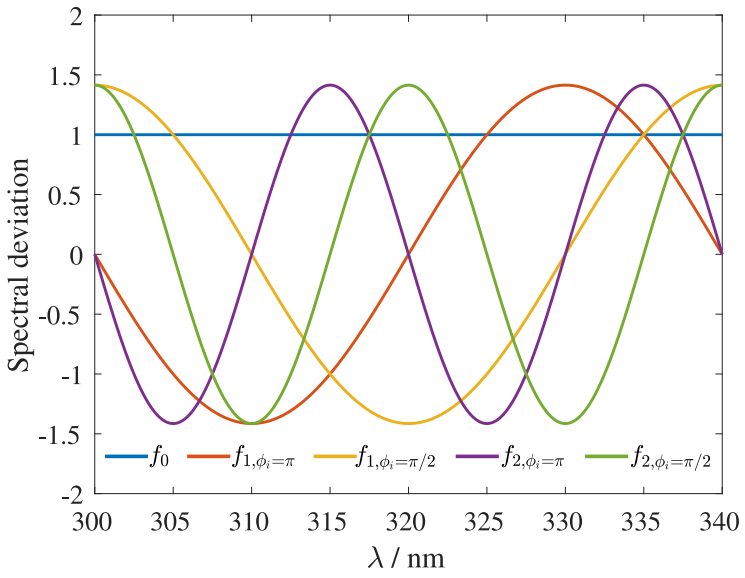


Figure 4. First three base functions with unity variance, f_1 and f_2 plotted with the phases $\phi_i = \pi$ and $\phi_i = \pi/2$.

The ~~obtained~~ deviation~~error~~ functions obtained with the cumulative Fourier series are used to distort the measured spectra $E(\lambda)$ as

$$E_e(\lambda) = [1 + \delta(\lambda)\, u(\lambda)]\, E(\lambda)\,, \tag{13}$$

where $u(\lambda)$ is the relative standard uncertainty of the spectrum. The corresponding spectral deviation is applied separately to the factors of Eq. (4), i.e., the extra-terrestrial solar spectrum, ozone absorption cross-section, and Rayleigh scattering.

Variances of the TOC values obtained by varying the weights γ_i and the phase terms ϕ_i give the desired uncertainties. Figure 5(a) presents how the uncertainty induced by deviation in spectral irradiance $E(\lambda)$ (circles) changes with increasing N. Each standard uncertainty of TOC in Fig. 5(a) was estimated from the MC results obtained by running the TOC retrieval 1000 times so that the phases ϕ_i and the weights γ_i of the base functions were independent at each round. Retrieved TOC deviations resemble the Gaussian distribution when the order of complexity of the deviation function is $N \geq 2$ as illustrated in Figs. 5(b) and 5(c). As we can see, *full* correlation with the base function $f_0(\lambda)$ at $N = 0$ causes a negligible~~small~~ uncertainty to TOC. The maximum at $N = 1$ gives uncertainty for an *unfavourable* case of correlations with base functions $f_0(\lambda)$ and $f_1(\lambda)$. Cases $N = 80$ for QASUME, $N = 100$ for BTS, or $N = 125$ for AVODOR correspond to the Nyquist criterion $(N = n/2)$ with base functions and give the uncertainty with no spectral correlations (only *random* noise). The obtained TOC value is affected most by spectral distortion that mimics the spectral shape of the ozone absorption ~~that may be approximated by a slope~~. The first combination of constant offset and one sinusoidal function with two sign changes within the region of interest is closest to this extreme.

The ozone absorption cross-section $\alpha_{O_3}(\lambda, T_{\text{eff}})$ is a direct multiplier of TOC, and thus the uncertainty in TOC is directly proportional to the deviations in the ozone absorption cross-section. The uncertainty in TOC arising from the spectral deviation

in $\alpha_{\mathrm{O_3}}(\lambda, T_{\mathrm{eff}})$ is plotted as crosses in Fig. 5(a) as a function of the increasing N. The ozone absorption cross-section, Serdyuchenko–Gorshelev data set (Weber *et al.* (2016)), has a wavelength step size of 0.01 nm, and thus the standard uncertainty of 0.05% at the Nyquist criterion $N = 2500$ is outside the range displayed in Fig. 5(a). Unlike the negligible effect of *full* spectral correlation in the spectral irradiance $E(\lambda)$ in TOC, *full* correlation ($N = 0$) in the ozone absorption cross-section produces approximately the same uncertainty as *unfavourable* correlation ($N = 1$). Generally, these results cannot be known before the analysis is carried out, using a method that does not have any internal limitation to the shape of the deviation~~error~~ function $\delta(\lambda)$. In some other cases, the uncertainty extreme appears at higher N-values, e.g., $N = 3$ noted for correlated colour temperature by ~~in (~~Kärhä *et al.* (2017b)~~)~~.

One major problem in uncertainty estimation is that typically many of the correlations in spectral irradiance data are unknown. Figure 5(a) can be used to find limits for the uncertainties assuming different correlation scenarios. In the analysis carried out in this paper, we estimate for each uncertainty component which kind of correlations it most likely has. For this, we divide the correlation into three categories, *full*, *unfavourable*, and *random* and estimate fractions on the assortment of these correlations. *Full* indicates that relative deviation~~error~~ is wavelength independent, such as with distance setting in spectral irradiance measurements~~geometrical factors~~. *Random* indicates no correlation between spectral values~~(noise)~~. As can be seen in Fig. 5(a), the uncertainty caused by noise depends on the Nyquist criterion, increasing with the smaller number of base functions. *Unfavourable* indicates an unknown deviation with systematic spectral structure~~spectrally varying error such as a slope~~ that produces a large deviation~~error~~ in TOC.

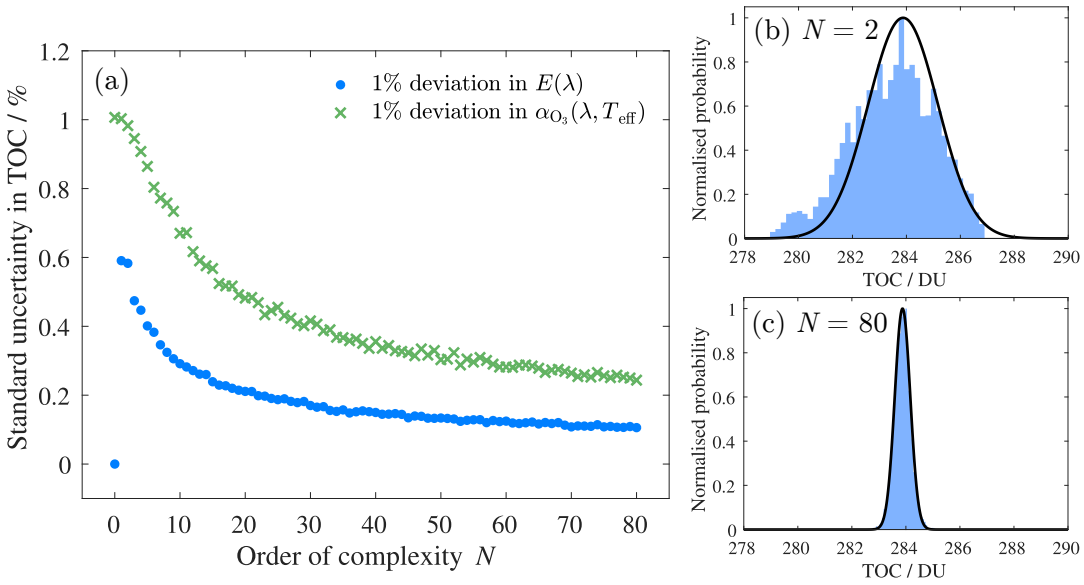


Figure 5. (a) Standard u~~U~~ncertainties of TOC at local noon as a function of the order of complexity N for QASUME spectroradiometer with 1% deviation in spectral irradiance $E(\lambda)$ plotted as circles and 1% deviation in ozone absorption cross-section $\alpha_{O_3}(\lambda, T_{\text{eff}})$ plotted as crosses. The distributions of TOC values arising from 1% deviation in $E(\lambda)$ with the order of complexity of $N = 2$ in (b) and $N = 80$ in (c). The black solid curve denotes Gaussian distribution. ~~at three different levels of uncertainty in spectral irradiance $E(\lambda)$ at 295 nm – 340 nm with the wavelength interval of 0.5 nm indicated with symbols in the figure legend. The black solid lines obtained by multiplying the 1% uncertainties indicate the scalability of the model (Kärhä et al. (2017a)).~~

4.2 Uncertainty budgets for spectroradiometers

Uncertainty budgets of the direct solar spectral irradiance measurements for QASUME, BTS, and AVODOR are presented in Tables 1–3. The tables also state fractions that we estimate for the three correlation types introduced for each component. The uncertainties due to radiometric calibration include factors such as the uncertainty of the standard lamp used, and the additional uncertainty due to noise and alignment. QASUME has been validated using various methods, thus the uncertainty due to calibration is low, 0.55% (Hülsen *et al.* (2016)). For QASUME and BTS, we assume the correlations to be equally distributed between *full* correlation, *unfavourable* correlation, and *random* correlation (Kärhä *et al.* (2018)). Spectra measured with AVODOR are significantly noisier, thus half of the calibration uncertainty is associated to the random component. Values for instability of the calibration lamp are based on long-term monitoring. The lamp irradiances have been noted to gradually drop with no significant wavelength structure within the wavelength region concerned. Non-linearity values are estimations of the operators of the devices. Non-linearity is typically manifested so that the responsivity of the device changes gradually from high readings to low readings. This can cause significant change in the TOC values, thus we assume the correlation type to be *unfavourable*. Uncertainties due to device stability and temperature dependence are based on long-term monitoring. The changes have been found to be independent on wavelength in the region concerned, thus *full* correlation is assumed. Noise

is the average standard deviation of typical measurements at noon over the wavelength region concerned. The wavelength scales of the devices have been checked using emission lines of gas discharge lamps. The uncertainty values given are the estimated standard deviations of the possible remaining errors after corrections. Wavelength error can introduce a significant change in TOC, because it introduces an error in the form of the derivative of the spectral irradiance. Thus, *unfavourable* correlation is assumed. Most of the uncertainty components are slightly wavelength dependent but to simplify simulations, average uncertainty values are used over the wavelength range between 300 nm and 340 nm.

Table 1. Uncertainties of the measurement for QASUME spectroradiometer.

QASUME	Standard		Correlation	
Source of uncertainty	uncertainty	*full*	*unfavourable*	*random*
in measured $E(\lambda)$	%		Fraction	
Radiometric calibration	0.55	$1/\sqrt{3}$	$1/\sqrt{3}$	$1/\sqrt{3}$
250 W lamp stability	0.14	1	0	0
Non-linearity	0.25	0	1	0
Stability	0.60	1	0	0
Temperature dependence	0.20	1	0	0
Measurement noise	0.20	0	0	1
Wavelength shift	0.10	0	1	0
Combined uncertainty ($k = 1$)	0.91%	0.72%	0.42%	0.38%

Table 2. Uncertainties of the measurement for BTS spectroradiometer (Zuber *et al.* (2017b)).

BTS	Standard		Correlation	
Source of uncertainty	uncertainty	*full*	*unfavourable*	*random*
in measured $E(\lambda)$	%		Fraction	
Radiometric calibration	0.80	$1/\sqrt{3}$	$1/\sqrt{3}$	$1/\sqrt{3}$
250 W lamp stability	0.20	1	0	0
Non-linearity	0.40	0	1	0
Stability	0.80	1	0	0
Temperature dependence	0.10	1	0	0
Measurement noise	0.20	0	0	1
Wavelength shift	0.10	0	1	0
Combined uncertainty ($k = 1$)	1.24%	0.95%	0.62%	0.50%

Table 3. Uncertainties of the measurement for AVODOR spectroradiometer.

AVODOR	Standard	Correlation		
Source of uncertainty	uncertainty	*full*	*unfavourable*	*random*
in measured $E(\lambda)$	%		Fraction	
Radiometric calibration	2.50	1/2	1/2	$1/\sqrt{2}$
250 W lamp stability	0.14	1	0	0
Non-linearity	0.50	0	1	0
Stability	0.60	1	0	0
Temperature dependence	0.20	1	0	0
Measurement noise	1.30	0	0	1
Wavelength shift	0.10	0	1	0
Combined uncertainty ($k = 1$)	2.94%	1.41%	1.35%	2.19%

4.3 Uncertainty budget for total ozone column

Table 4 presents an uncertainty budget for TOC that would be obtained with the high-accuracy QASUME spectroradiometer at local noon. All major uncertainty components are listed and estimated. The uncertainty components ~~stating fractions~~divided to the three correlation types have been analysed with the new model. The other components (a)–(d) in Table 4 have been solved using traditional MC modelling~~analytically~~ because the mechanism for the uncertainty propagating to TOC is known.

Table 4. An example uncertainty budget for QASUME spectroradiometer measured at local noon on the clear day of 17 September 2016. The last column states the standard deviations in $u(\mathrm{TOC})$ corresponding to each individual uncertainty component~~, when measuring~~ for TOC = 284~~2~~ DU retrieved from the QASUME spectrum using ~~with~~ the spectral range of 300 nm – 340 nm ~~and~~ at the solar zenith angle of 26.35°. The expanded uncertainty stated, $U(\mathrm{TOC})$ =3.6~~7~~ DU, has been obtained by multiplying the combined standard uncertainty with a coverage factor $k = 2$.

Source of uncertainty	Standard uncertainty in $E(\lambda)$ %	Standard uncertainty in exponent %	Correlation *full*	Correlation *unfavourable*	Correlation *random*	$u(\mathrm{TOC})$ DU
			Fraction			
Measurement						
Radiometric calibration	0.55		$1/\sqrt{3}$	$1/\sqrt{3}$	$1/\sqrt{3}$	0.44~~3~~
250 W lamp stability (one year)	0.14		1	0	0	0.00~~2~~
Non-linearity	0.25		0	1	0	0.33~~5~~
Stability	0.60		1	0	0	0.00~~10~~
Temperature dependence	0.20		1	0	0	0.00~~3~~
Measurement noise	0.20		0	0	1	0.06~~7~~
Wavelength shift	0.10		0	1	0	0.13~~4~~
Uncertainties related to $E(\lambda)$						
Extra-terrestrial spectrum (Gröbner *et al.* (2017))	1.00		$1/\sqrt{3}$	$1/\sqrt{3}$	$1/\sqrt{3}$	0.95~~1.00~~
Uncertainties related to exponent of Eq. (4)						
O_3 cross-section (Weber *et al.* (2016))		1.5	0.23	0.23	0.95	1.41
Rayleigh scattering (Bodhaine *et al.* (1999))		0.1	$1/\sqrt{3}$	$1/\sqrt{3}$	$1/\sqrt{3}$	0.09
O_3 layer altitude of 26 km, $u = 0.5$ km		(a)				0.01
Rayleigh layer altitude of 5 km, $u = 0.5$ km		(b)				0.00
Temperature of O_3 cross-section at 228 K, $u = 1$ K		(c)				0.28
Station pressure of 772.8 hPa, $u = 1.3$ hPa		(d)				0.05
					$U(\mathrm{TOC})$	**3.6~~70~~**

(a) Air mass m_{TOC} varies as a function of the altitude of O_3 layer.

(b) Air mass m_{R} varies as a function of the altitude of Rayleigh scattering layer.

(c) O_3 cross-section varies as a function of temperature.

(d) Rayleigh scattering depends on the station pressure.

The uncertainties produced in TOC were obtained separately for all components, by setting other uncertainties to zero. Division of the correlation to the three categories introduced are stated for each row as fractions r_{full}, r_{unfav}, and r_{random}. For

example, the ground-based spectrum $E(\lambda)$ measured is deviated with the three correlation components as

$$E_{\mathrm{e}}(\lambda) = (1 + u\, r_{full} f_0(\lambda)) \cdot \left(1 + u\, r_{unfav} \sum_{i=0}^{1} \gamma_i' f_i(\lambda)\right) \cdot \left(1 + u\, r_{random} \sum_{j=0}^{N} \gamma_j'' f_j(\lambda)\right) E(\lambda)\,, \tag{14}$$

where ~~u is the standard uncertainty of the component and γ_0,~~ γ_i', and γ_j'' are independent MC variables generated using Eq. (12).

It is worth noting that not all uncertainty components affect the spectrum $E(\lambda)$ directly, but via the exponent of Eq. (4). Corresponding formulas are used to evaluate the effect of uncertainties in extra-terrestrial solar spectrum, ozone absorption cross-section, and Rayleigh scattering. The last column in Table 4 states the standard uncertainties produced by each uncertainty component with the assumed fractions, calculated with an irradiance spectrum measured at local noon with QASUME (Hülsen *et al.* (2016)). The expanded uncertainty of the TOC, obtained as the square sum of the individual components and multiplied with coverage factor $k = 2$, for this spectral measurement is 3.6~~7~~ DU (1.3%).

The QASUME spectroradiometer has a combined measurement standard uncertainty of 0.91% (Hülsen *et al.* (2016)) arising from the uncertainty components explained in Section 4.2. The uncertainty components stated are typical in solar UV spectral irradiance measurements (Bernhard and Seckmeyer (1999)). Division of the radiometric calibration uncertainty to equal fractions of $1/\sqrt{3}$ is based on typical spectral correlations noted in intercomparisons (Kärhä *et al.* (2018)). The lamp data obtained from national standard laboratories are highly accurate but also typically spectrally correlated. Due to very low noise, elements such as interpolation functions, offsets and slopes are present in the data. When the calibration is transferred further, uncertainty increases due to noise, and correlations reduce. We thus assume that in this high accuracy calibration, there are equal amounts of *fully* correlated, *unfavourably* correlated, and *uncorrelated* uncertainties included. ~~The standard lamp used has been noted to decrease in intensity with no spectral change within the region of interest, thus the resulting uncertainty is fully correlated. Non-linearity may cause spectral change, as low values will be systematically distorted with respect to higher values, thus unfavourable correlation is assumed. Stability of the QASUME readings during a day has been noted to be wavelength independent, so is the effect of temperature. Noise contains no correlations, whereas a wavelength shift will introduce unfavourable correlations.~~

For $E_{\mathrm{ext}}(\lambda)$, we use QASUME-FTS (Gröbner *et al.* (2017)). We assume the correlation to be similar to a standard lamp, thus containing equal fractions of *full*, *unfavourable*, and *random* correlations. The QASUME-FTS is provided in air wavelengths with a step size of 0.01 nm. Otherwise, the wavelength shift due to vacuum-air interface should be corrected from~~for~~ the extra-terrestrial spectrum.

As the reference ozone absorption cross-section, the Serdyuchenko–Gorshelev ~~(SG)~~ data set given in air wavelengths with a step size of 0.01 nm, was used with 1.5% standard uncertainty (Weber *et al.* (2016)). The systematic and random uncertainties of the Serdyuchenko–Gorshelev~~SG~~ data set are given separately (Weber *et al.* (2016)). We further estimate that the systematic uncertainty given may include equal fractions of *fully* correlated and *unfavourably* correlated uncertainty. Thus, according to that we use the fractions of 0.23 for *full*, 0.23 for *unfavourable* and 0.95 for *random* correlations. *Full* correlation with a fraction of 0.23 produces a standard uncertainty of 0.96 DU, *unfavourable* correlation with a fraction of 0.23 produces a standard uncertainty of 1.01 DU and *random* correlation with a fraction of 0.95 produces a standard uncertainty of 0.22 DU.

AltogetherAs can be seen, the ozone cross-section causes an uncertainty of 1.41 DU to TOC, and is thus the dominating component in the uncertainty. If the fractions of correlations are not equally distributed between *full* and *unfavourable*, uncertainty in TOC does not change significantly from 1.41 DU. Fractions of 0.31 for *full*, 0 for *unfavourable* (or fractions of 0 for *full*, 0.31 for *unfavourable*), and 0.95 for *random* correlations, cause an uncertainty of 1.33 DU (or 1.43 DU) in TOC. Fractions of 0 for full, 0.31 for unfavourable, and 0.95 for random correlations, cause an uncertainty of 1.43 DU in TOC.

For components (a)–(d) in Table 4, the mechanism of contributing to the uncertainty of TOC is known. We know the standard uncertainty of the O_3 layer altitude of 26 km to be $u = 0.5$ km, so we vary the altitude accordingly and note the variance of the resulting TOC. Rayleigh scattering and aerosols are set at the altitude of 5 km ± 0.5 km, which influences the relative air mass $m_{\mathrm{R}} \approx m_{\mathrm{AOD}}$ (Gröbner *et al.* (2017)Gröbner and Kerr (2001)). This component causes negligible uncertainty to TOC. For calculating $\tau_{\mathrm{R}}(\lambda, P_0, z_0, \phi)$, we use a model by (Bodhaine *et al.* (1999)) with 0.1% uncertainty with equal estimated fractions of correlation types. The correlation has been obtained by studying how this data deviates from the model by (Nicolet (1984)). The ozone and temperature profiles were measured with a sonde during the campaign and based on the profiles the effective altitude of the ozone layer was at 26 km $\pm$ 0.5 km at the effective temperature of 228 K $\pm$ 1 K. The effect on TOC was obtained by randomly varying the altitude with the Gaussian uncertainty distribution. The same applies to air pressure that was 772.8 hPa with a standard uncertainty of 1.3 hPa. The effect of temperature on TOC was obtained by randomly varying the temperature with its standard uncertainty of 1 K. Varying the temperature systematically changes the spectral ozone absorption cross-section according to Eq. (6).

Stray light that affects TOC results at large solar zenith angles (Appendix A) has not been accounted for in the uncertainty analysis due to lack of information. Proper correction and estimation of the uncertainty due to stray light would require the stray light correction matrix to be measured. The effect of stray light, typical for measurements with array spectroradiometers, was reduced from TOC results by fitting BTS and AVODOR spectra with the OLS method. Then, these TOC values were corrected to be compatible with the RLS results by Eq. (9). The correction factor involves standard uncertainty that is 0.1% (0.28 DU) for BTS and 1.1% (3.10 DU) for AVODOR.

5 Results and discussion

The calculated TOC values obtained by the three different spectroradiometers on 17 September 2016 are presented in Fig. 6. Expanded uncertainties of the TOC values calculated are stated in DU as error bars. Measurement results of Brewer spectrophotometer #183 used as a reference in the intercomparison have been included in Fig. 6 as well. Looking at the absolute values of TOC in Fig. 6, we may conclude that the results of QASUME and Brewer #183 are in excellent agreement. Also the TOC values estimated for BTS and AVODOR are in agreement with Brewer #183 within uncertainties.

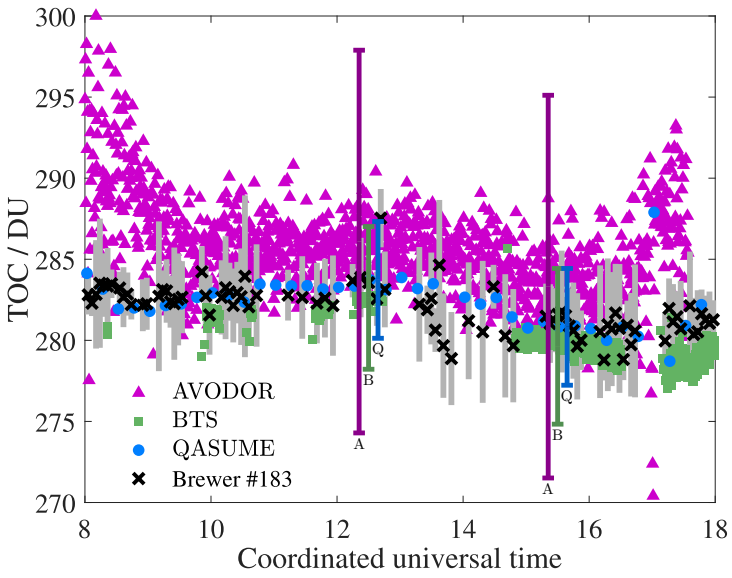


Figure 6. ~~Derived t~~Total ozone columns (TOC) derived from the solar UV spectra from 300 nm to 340 nm with expanded uncertainty bars ($k = 2$) for QASUME indicated as blue circles~~(300 nm – 340 nm, noise floor of 10^{-5})~~, BTS indicated as green squares~~(300 nm – 340 nm, noise floor of $5 \cdot 10^{-3}$)~~, and AVODOR indicated as magenta triangles~~(305 nm – 340 nm, noise floor of $5 \cdot 10^{-3}$)~~. The TOC measured with the Brewer #183 is plotted as black crosses with grey ~~thick~~ uncertainty bars ($k = 2$).

~~The uncertainty budget used for QASUME was presented in Table 4. For AVODOR and BTS, the components related with measurement were 2.94% and 1.24% as shown in Tables 2 and 3. For BTS, the uncertainty was distributed between correlations as follows: 0.95% for full, 0.62% for unfavourable, and 0.50% for random correlations. For AVODOR, the uncertainty was distributed as 1.41% for full, 1.35% for unfavourable, and 2.19% for random correlations. The uncertainty due to calibration was 0.8% 5 for BTS (Zuber et al. (2017a)) and 2.5% for AVODOR. AVODOR has problems with noise, thus, half of the uncertainty was assumed to be uncorrelated. Non-linearity was assumed to be 0.5% for AVODOR and 0.4% for BTS. Measurement noise introduced uncertainties of 1.3% and 0.2% to AVODOR and BTS, respectively. With BTS, stability of the lamp was 0.2%, stability of the device was 0.8% and temperature sensitivity was 0.1%.~~

The relative uncertainties of the TOC values obtained with the three instruments are shown in Fig. 7. The expanded uncertainties of the TOC data sets at local noon are 3.6~~7~~ DU (1.3%) for the QASUME spectroradiometer, 4.4~~3.7~~ DU (1.5~~3~~%) for the BTS spectroradiometer, and 13.3~~9.3~~ DU (4.7~~3.3~~%) for the AVODOR spectroradiometer. The expanded uncertainties stated include the standard deviations of 0.28 DU for BTS and 3.10 DU for AVODOR, arising from the fitting method used. ~~The uncertainties are generally higher at noon, as the uncertainty of measurement influences more with lower air mass.~~

It is of interest to compare the obtained uncertainties with values assuming no correlations. If we neglect correlations, i.e., we assume the fractions in Table 4 to be 0 for the *full* and *unfavourable* correlations and 1 for the *random* correlation, and run the simulations with the spectrum measured at local noon, we obtain the expanded uncertainty $U_{\mathrm{QASUME}}(\mathrm{TOC}) = 0.9$~~1.1~~ DU

(0.34%), $U_{\mathrm{BTS}}(\mathrm{TOC}) = 1.1$ DU (0.4%), and $U_{\mathrm{AVODOR}}(\mathrm{TOC}) = 7.73.3$ DU (2.71.2%). These values are on average a factor of 3 lower than the uncertainties accounting for correlations. This analysis assumes random noise only.

A typical practice in an analysis like this is to add a component introduced by the standard deviation of the fit to the uncertainty. The standard uncertainty to be added to $u(\mathrm{TOC})$the TOC because of theis standard deviation of the fit is 0.28 DU with QASUME, 0.74 DU with BTS and 3.42.45 DU with AVODOR, raising the corresponding total expanded uncertainties to 1.03 DU (0.34%), 1.8 DU (0.67%), and 10.35.9 DU (3.62.1%). The results are generally in agreement within these lower uncertainties as well. However, comparing differences in the TOC results of the different spectroradiometers does not represent the uncertainty in the absolute TOC scale, since the ozone retrieval algorithm uses the same extra-terrestrial spectrum and ozone absorption cross-section for all the instruments. Changing the extra-terrestrial spectrum or the ozone absorption cross-section to another data set may shift all the TOC values of the instruments beyond the latter low uncertainties.

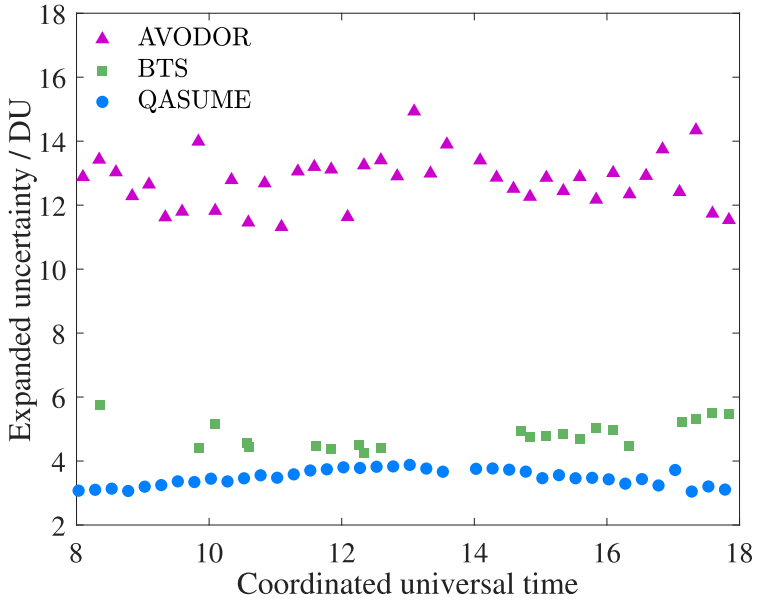


Figure 7. Relative expanded uncertainties of the total ozone columns derived from QASUME, BTS, and AVODOR spectra obtained by the relative least squares fitting method.

Figure 6 shows that the effect of stray light can be effectively reduced from TOC results by using the OLS fitting described in Appendix A with the correction in Eq. (9). For example, BTS and QASUME results are in good agreement even at the largest zenith angles. TOC results of AVODOR are in agreement at noon, but the results before 09:00 and after 17:00 deviate from the other instruments by 10 DU.

Also the results of BTS are in good agreement at noon, but start deviating slightly at higher zenith angles. The ozone levels measured by the AVODOR spectroradiometer have a large discrepancy of 10 DU at noon and a strong dependence on the solar elevation angle. The offset of the AVODOR, and the angular structure of both spectrographs, are due to spectral correlations

present in the measurements. Stray light and non-linearity cause systematic effects in measurements, which most likely explain the dependence as a function of solar zenith angle.

The 10 DU deviation of the AVODOR at noon is large, barely within the calculated uncertainty, so it was studied in more detail. It appears that the reason is a systematic deviation either in the linearity, stray light properties, or the calibration of the device. Further information may be obtained by studying the AOD that is the other term to be fitted in addition to TOC, plotted for the three instruments in Fig. 7(a). The aerosol optical depth is a property of the atmosphere, and thus all the instruments should see the same AOD spectrum when the measurements are carried out at the same place and time. This applies more or less to QASUME and BTS, but for AVODOR $\tau_{\mathrm{AOD}}(\lambda)$ does not resemble AOD anymore.

Figure 7. Aerosol optical depth (AOD) spectra obtained from QASUME, BTS, and AVODOR data sets by fitting, using the Ångström exponent based AOD model of Eq. (4) in (a) and the linear AOD model of Eq. (13) in (b). The solid lines indicate AOD at noon. The dashed lines are for a later measurement with the larger zenith angle.

There is a systematic wavelength dependent error in the measured spectra and the modelling algorithm tries to minimise this with AOD. The phenomenon was further studied by changing the $\tau_{\mathrm{AOD}}(\lambda)$ of (Ångström (1964)) to a linear model as where λ is the wavelength in nm, and both a and b are fitting parameters in addition to TOC. This model is more parametrised so it can cause more changes to match with the measured spectra. Changing the AOD model had a drastic effect on the AVODOR results. The absolute TOC values dropped by several per cent, and were in much better agreement with the other devices. Also the dependence of the TOC on the zenith angle improved. However, $\tau_{\mathrm{AOD}}(\lambda)$ obtained did not describe the physical attenuation by aerosols anymore, which can be seen as the changing slopes in Fig. 7(b), indicating that the AOD part of the model was correcting for something different than AOD.

6 Conclusions

In this work, we introduced one possible way to take into account spectral correlations in the uncertainties of the atmospheric ozone retrieval and estimated the TOC uncertainties obtained from the spectral data of three different spectroradiometers, measured at the ATMOZ field measurement~~Izaña~~ campaign at the Izaña Atmospheric Observatory. It should be noted that the method proposed has a drawback that the unknown correlations have to be approximated based on experience. However, the method has merits in estimating the order of magnitude of possible uncertainties accounting for correlations. The typical assumption made, that uncertainties are spectrally uncorrelated, is just an assumption as well, and in many cases not valid. The uncertainty values obtained with the new model are higher than the uncertainties obtained with the traditional method neglecting correlations because some of the major uncertainty components may contain systematic spectral deviations. These results demonstrate the importance of accounting for correlations. If their origins and magnitudes are known, they can be accounted for precisely using methods of (JCGM 100 (2008)).

The behaviour of the AVODOR spectroradiometer is a good example about what spectral correlations in uncertainties due to systematic error components may cause. With AVODOR, the results deviate 10 DU from the other instruments, which is more than the uncertainties assuming no correlations predict, 5.9 DU. Our approach gives uncertainty of 9.3 DU which barely

~~covers the noted deviation. It was also noted that the modelling of AOD does not work with this device, because the part of the model meant for correcting AOD starts to correct for the systematic error in the measurement results, and thus affects the TOC obtained. Spectroradiometers to be used for deriving TOC should be better with respect to systematic effects. Thus, they should be corrected for stray light and linearity, and care should be taken in their calibration not to leave systematic errors in their responsivity. The QASUME and BTS spectroradiometers are significantly better in this respect.~~

The new model uses ~~the~~ similar approach to our previously developed MC uncertainty model for correlated colour temperature (CCT) (Kärhä *et al.* (2017b)). In the article, we demonstrated the use of the model for calculating the CCT of a Standard Illuminant A. For Standard Illuminant A, the case representing uncertainty with *unfavourable* correlations in CCT was found at $N = 3$. On the contrary, for the ozone retrieval the deviation at $N = 1$ ~~base functions~~ produces the largest uncertainty, which is in a way trivial compared with CCT. The use of a set of sine functions as base functions was originally developed for the more complicated situation of CCT where it was not known where the *unfavourable* uncertainty would be. When we now have analysed the situation, an uncertainty arising from~~the case of~~ unfavourable correlations in the ozone retrieval could as well be modelled e.g. using a combination of *full* spectral deviation,~~the derivative of the solar UV irradiance, a derivative of the ozone cross-section, or~~ a simple slope, and a parabola as the deviation function mimicking *unfavourable* correlations~~, which would simplify the model~~. This is discussed in more detail in Appendix B.

The new MC method for estimating uncertainties in TOC in the presence of systematic spectral deviations provides more complete estimations of the uncertainty budget compared with the traditionally used methods. The TOC values retrieved from different instrument data were well in agreement within the uncertainties estimated with the new MC method. Although the TOC results obtained using different instruments have good agreement, these differences do not represent the uncertainty of the absolute TOC scale. The TOC uncertainties we have estimated cover also possible offsets in the absolute TOC scale, arising from the uncertainties in the ozone absorption cross-section and extra-terrestrial spectrum that are the dominating components in the uncertainty budget.

Appendix A: Selecting the least squares fitting method

Two of the instruments being compared suffer from stray light and noise that distort the TOC results. In earlier studies e.g. by Herman *et al.* (2015), stray light in array spectroradiometers has been noted to decrease TOC values at large solar zenith angles resulting in an inverted U-shape dependence of the diurnal TOC variation. The effect of stray light can be compensated by reducing the effect of short-wavelength tail, either by limiting the wavelength range or the dynamic range used, or by weighting the results. We studied various weightings and selection methods for data in order to find an objective way to perform the ozone retrieval method and to analyse the results.

Figure A1 shows the TOC results as a function of time analysed with three different weighting methods for least squares minimisation. The methods include a relative least squares fitting (RLS) in Eq. (8) with $w(\lambda) = E(\lambda)^{-2}$, RLS fitting with the dynamic range (DR) limited to avoid stray light, and an ordinary least squares fitting method (OLS) with $w(\lambda) = 1$.

TOC values estimated for QASUME in Fig. A1(a) have no significant solar zenith angle dependence, and limiting the dynamic range of RLS fitting only affects individual TOC values in the early morning and late afternoon compared with RLS fitting over the complete spectral range. Using OLS fitting method, the diurnal variation of the TOC remains, but the values are underestimated by a constant factor of 1.013.

TOC values estimated for BTS in Fig. A1(b) and AVODOR in Fig. A1(c) have severe dependence on the solar zenith angle when using RLS fitting method. The solar zenith angle dependence decreases to approximately half when limiting the dynamic range to exclude the baseline noise from the fitting. Best results are obtained by using the OLS minimisation which practically removes the solar zenith angle dependence for both BTS and AVODOR but introduces an offset to TOC results. Almost similar offset in TOC results was noted for QASUME. The OLS method violates against the heteroskedasticity of our data sets, since we know that the absolute spectral uncertainties are not constant at the wavelength range studied. The only reason to use the OLS method for array spectroradiometers is to reduce the effect of stray light from TOC results. Hence, we correct the TOC results obtained with the OLS method with correction factors estimated from the ratios of TOC values determined from the local noon spectra using both RLS and OLS methods. The correction factors were averaged over 10 samples around noon, being 1.006 for BTS and 1.013 for AVODOR with standard deviations of 0.1% (0.28 DU) and 1.1% (3.10 DU), respectively. This correction makes the TOC results comparable with devices analysed using the RLS method.

For QASUME, we use RLS minimisation with the dynamic range limited to four orders of magnitude, as it uses most of the useful undistorted data. This is also consistent with the earlier methods used with monochromator-based spectroradiometers e.g. by Huber *et al.* (1995).

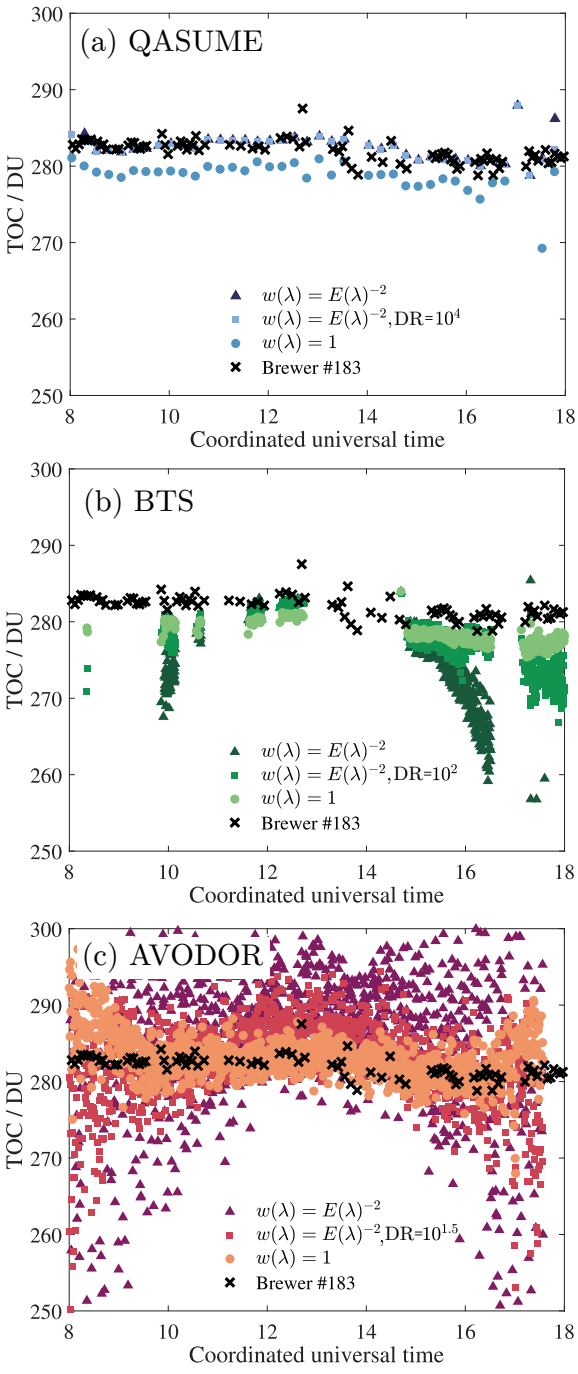


Figure A1. TOC values during the day estimated using different weightings in the least squares minimisation for QASUME (a), BTS (b) and AVODOR (c). TOC values for Brewer #183 are plotted as black crosses for comparison. The colour codes and the associated figure legends denote the weighting and the dynamic range (DR) used.

Appendix B: TOC uncertainties obtained using Chebyshev polynomials

In the MC uncertainty analysis, it is possible to use other orthogonal sets of functions instead of sinusoidal functions, such as Chebyshev polynomials shown in Fig. B1. A Chebyshev polynomial of the first kind $T_j(\lambda)$ of order j (Kreyszig (2006), p. 209) is defined as

$$T_j(\lambda) = \cos\left[j \arccos\left(\frac{2\lambda - \lambda_1 - \lambda_2}{\lambda_2 - \lambda_1}\right)\right], \tag{B1}$$

where λ_1 is the short wavelength limit and λ_2 is the long wavelength limit for the spectra measured. To create arbitrary spectral deviations with unity variance, each Chebyshev polynomial, except for $g_0(\lambda) = T_0(\lambda) = 1$, needs to be normalised to unity variance as

$$g_j(\lambda) = \frac{T_j(\lambda)}{\sigma_j}, \tag{B2}$$

where σ_j is the standard deviation of $T_j(\lambda)$. In practice, Chebyshev polynomials in Fig. (B1) can be generated fast using recurrence (e.g. Fateman (1989)) as

$$T_j(\lambda) = 2\left(\frac{2\lambda - \lambda_1 - \lambda_2}{\lambda_2 - \lambda_1}\right) T_{j-1}(\lambda) - T_{j-2}(\lambda), \tag{B3}$$

where $T_0(\lambda) = 1$, and $T_1(\lambda) = (2\lambda - \lambda_1 - \lambda_2)/(\lambda_2 - \lambda_1)$ is a straight line. The scaling by the standard deviation according to Eq. (B2) is performed after generating the Chebyshev polynomials.

Each base function of the cumulative Fourier series in Eq. (10) was formed with sine (odd) and cosine (even) terms (Kreyszig (2006), p. 628),

$$\sin\left[i\left(2\pi\frac{\lambda - \lambda_1}{\lambda_2 - \lambda_1}\right) + \phi_i\right] = \cos(\phi_i) \cdot \sin\left(2\pi i \frac{\lambda - \lambda_1}{\lambda_2 - \lambda_1}\right) + \sin(\phi_i) \cdot \cos\left(2\pi i \frac{\lambda - \lambda_1}{\lambda_2 - \lambda_1}\right), \tag{B4}$$

where the phase ϕ_i is an equally distributed MC variable between 0 and 2π. Hence, the analysis based on Chebyshev polynomials to be compatible with the sinusoidal approach, each base function $f_i(\lambda)$ with index $i \geq 1$ is formed by the combination of odd $(2i-1)$ and even $(2i)$ terms as

$$f_i(\lambda) = \cos(\phi_i) \cdot g_{2i-1}(\lambda) + \sin(\phi_i) \cdot g_{2i}(\lambda), \tag{B5}$$

where the weights $\cos(\phi_i)$ and $\sin(\phi_i)$ set the variance of $f_i(\lambda)$ to unity. These base functions $f_i(\lambda)$ can be used with Eqs. (10) and (13).

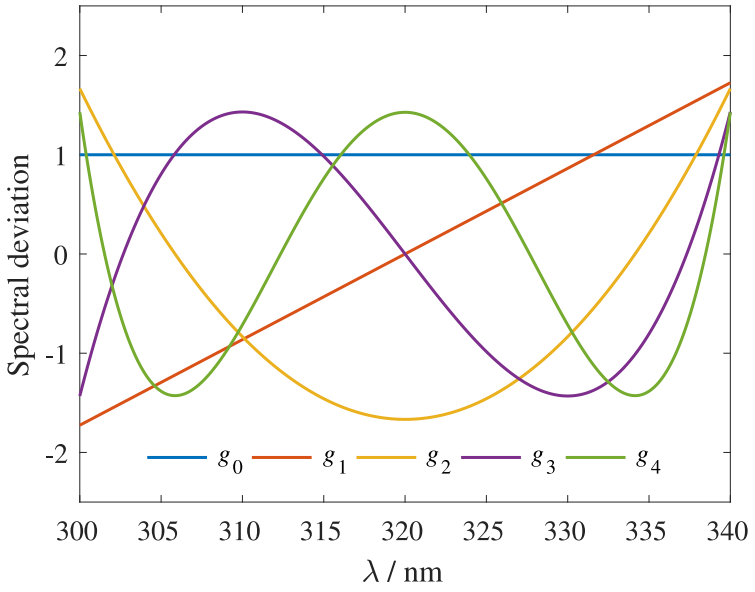


Figure B1. First five Chebyshev polynomials with unity variance corresponding to Fig. 4 with sinusoidal base functions.

Figure B2 compares standard uncertainties of TOC obtained by generating arbitrary spectral deviations using Chebyshev polynomials as base functions (circles and triangles), formed using Eq. (B5), with those obtained by using sinusoidal base functions (crosses). The uncertainties change slightly at the lowest complexity orders of deviations when sinusoidal base functions are replaced with Chebyshev polynomials, but essentially the results are similar.

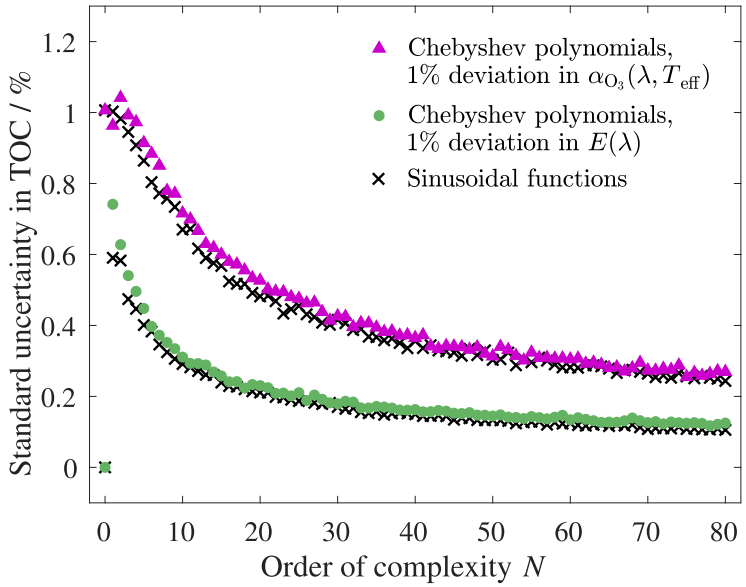


Figure B2. Standard uncertainties of TOC simulated using a local noon spectrum of QASUME with Chebyshev polynomials as base functions. The TOC uncertainties with the input standard deviation of 1% in the spectral irradiance values are shown as green circles and the TOC uncertainties with the input standard deviation of 1% in the ozone absorption cross-section are shown as magenta triangles. Standard uncertainties simulated with sinusoidal base functions (black crosses) from Fig. 5(a) are plotted for comparison.

Acknowledgements. Peter Sperfeld from PTB is acknowledged for measuring and providing the BTS data set. Alberto Redondas and all personnel from Izaña Atmospheric Research Center, AEMET, Tenerife, Canary Islands, Spain, are acknowledged for measuring and providing the Brewer #183 data set and the environmental parameters such as the sonde data. A.V. is grateful for the incentive grant by the Emil Aaltonen Foundation, Finland. This work has been supported by the European Metrology Research Programme (EMRP) within the joint research project ENV59 “Traceability for atmospheric total column ozone” (ATMOZ). The EMRP is jointly funded by the EMRP participating countries within EURAMET and the European Union.

References

ATMOZ field measurement campaign 2016,
URL = http://rbcce.aemet.es/2015/11/24/atmoz-intercomparison-campaign-at-izana-tenerife-september-2016/

Ångström A.: The parameters of atmospheric turbidity, Tellus, 16, 64–75, 1964.

Ball W. T., Alsing J., Mortlock D. J., Staehelin J., Haigh J. D., Peter T., Tummon F., Stübi R., Stenke A., Anderson J., Bourassa A., Davis S. M., Degenstein D., Frith S., Froidevaux L., Roth C., Sofieva V., Wang R., Wild J., Yu P., Ziemke J. R., and Rozanov E. V.: Evidence for a continuous decline in lower stratospheric ozone offsetting ozone layer recovery, Atmos. Chem. Phys., 18, 1379–1394, 2018.

Beer A.: Bestimmung der Absorption des rothen Lichts in farbigen Flüssigkeiten, Annalen der Physik und Chemie, 86, 78–88, 1852.

Bernhard G. and Seckmeyer G.: Uncertainty of measurements of spectral solar UV irradiance, J. Geophys. Res., 104, 14321–14345, 1999.

Bodhaine B. A., Wood N. B., Dutton E. G., and Slusser J. R.: On Rayleigh Optical Depth Calculations, J. Atmos. Oceanic Tech., 16, 1854–1861, 1999.

Bouguer P.: Essai d'optique sur la gradation de la lumière (Paris, France: Claude Jombert, 1729) pp. 16–22.

ENV59 "Traceability for atmospheric total column ozone" (ATMOZ) 2014 – 2017, URL = http://projects.pmodwrc.ch/atmoz/

Fabry C. and Buisson H.: L'absorption de l'ultra-violet par l'ozone et la limite du spectre solaire, J. de Phys., 3, 196–206 (1913).

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.

Fateman R. J.: Lookup tables, recurrences and complexity, Proc. Int. Symp. Symbolic and Algebraic Computation (ISSAC '89), Portland, Oregon, USA, 17 – 19 July, 1989.

Fragkos K., Bais A. F., Balis D., Meleti C., and Koukouli M. E.: The Effect of Three Different Absorption Cross-Sections and their Temperature Dependence on Total Ozone Measured by a Mid-Latitude Brewer Spectrophotometer, Atmosphere-Ocean, 53, 19–28, 2015.

Gardiner B. G., Webb A. R., Bais A. F., Blumthaler M., Dirmhirn I., Forster P., Gillotay D., Henriksen K., Huber M., Kirsch P. J., Simon P. C., Svenoe T. , Weihs P., and Zerefos C. S.: European intercomparison of ultraviolet spectroradiometers, Environ. Technol., 14, 25–43, 1993.

GCOS-200: The Global Observing System for Climate: Implementation Needs (World Meteorological Organization, 2016).

Gröbner J. and Kerr J. B.: Ground-based determination of the spectral ultraviolet extraterrestrial solar irradiance: Providing a link between space-based and ground-based solar UV measurements, J. Geophys. Res., 106, 7211–7217, 2001.

Gröbner J., Kröger I., Egli L., Hülsen G., Riechelmann S., and Sperfeld P.: The high resolution extra-terrestrial solar spectrum determined from ground-based solar irradiance measurements, Atmos. Meas. Tech., 10, 3375–3383, 2017.

Gröbner J., Schreder J., Kazadzis S., Bais A. F., Blumthaler M., Görts P., Tax R., Koskela T., Seckmeyer G., Webb A. R., and Rembges D.: Traveling reference spectroradiometer for routine quality assurance of spectral solar ultraviolet irradiance measurements, Appl. Opt., 44, 5321–5331, 2005.

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.

Hicks J. S. and Wheeling R. F.: An efficient method for generating uniformly distributed points on the surface of an n-dimensional sphere, Commun. ACM, 2, 17–19, 1959.

Huber M., Blumthaler M., Ambach W., and Staehelin J.: Total atmospheric ozone determined from spectral measurements of direct UV irradiance, Geophys. Res. Lett., 22, 53–56, 1995.

Hülsen G., Gröbner J., Nevas S., Sperfeld P., Egli L., Porrovecchio G., and Smid M.: Traceability of solar UV measurements using the QASUME reference spectroradiometer, Appl. Opt., 55, 7265–7275, 2016.

JCGM 100, Evaluation of measurement data – Guide to the expression of uncertainty in measurement (ISO/IEC Guide 98–103), 2008.

Kerr J. B. and McElroy C. T.: Evidence for large upward trends of ultraviolet-B radiation, Science, 262, 1032–1034, 1993.

Kipp & Zonen: Brewer MkIII Spectrophotometer, Operator's manual Rev F, 2015.

Komhyr W. D., Mateer C. L., and Hudson R. D.: Effective Bass-Paur 1985 ozone absorption coefficients for use with Dobson ozone spectrophotometers, J. Geophys. Res., 98, 20451–20465, 1993.

Kreyszig E., *Advanced Engineering Mathematics*, 9th ed., John Wiley & Sons, 2006. pp. 209, 628. ISBN-13: 978-0-471-72897-9

Kärhä P., Vaskuri A., Gröbner J., Egli L., and Ikonen E.: Monte Carlo analysis of uncertainty of total atmospheric ozone derived from measured spectra, AIP Conference Proceedings, 1810, 110005, 2017a.

Kärhä P., Vaskuri A., Mäntynen H., Mikkonen N., and Ikonen E.: Method for estimating effects of unknown correlations in spectral irradiance data on uncertainties of spectrally integrated colorimetric quantities, Metrologia, 54, 524–534, 2017b.

Kärhä P., Vaskuri A., Pulli T., and Ikonen E.: Key comparison CCPR-K1.a as an interlaboratory comparison of correlated color temperature, J. Phys.: Conf. Ser. 972, 012012, 2018.

Lambert J. H.: Photometria sive de mensura et gradibus luminis, colorum et umbrae (Augsburg ("Augusta Vindelicorum"), Germany: Eberhardt Klett, 1760).

Levenberg K.: A method for the solution of certain non-linear problems in least squares, Quart. Appl. Math., 2, 164–168, 1944.

Manney G. L., Santee M. L., Rex M., Livesey N. J., Pitts M. C., Veefkind P., Nash E. R., Wohltmann I., Lehmann R., Froidevaux L., Poole L. R., Schoeberl M. R., Haffner D. P., Davies J., Dorokhov V., Gernandt H., Johnson B., Kivi R., Kyrö E., Larsen N., Levelt P. F., Makshtas A., McElroy C. T., Nakajima H., Concepción Parrondo M., Tarasick D. W., von der Gathen P., Walker K. A., and Zinoviev N. S.: Unprecedented Arctic ozone loss in 2011, Nature, 478, 469–475, 2011.

Marquardt D. W.: An Algorithm for Least-Squares Estimation of Nonlinear Parameters, J. Soc. Indust. Appl. Math., 11, 431–441, 1963.

McElroy C. T. and Fogal P. F., Ozone: From Discovery to Protection, Atmosphere-Ocean, 46, 1–13, 2008.

McKenzie R. L., Aucamp P. J., Bais A. F., Björn L. O. and Ilyas M.: Changes in biologically-active ultraviolet radiation reaching the Earth's surface, Photochem. Photobiol. Sci., 6, 218–231, 2007.

Meeus J.: *Astronomical Algorithms*, 2nd Edition, (Willmann-Bell, Richmond, VA, USA), 1998.

Molina M. J. and Rowland F. S.: Stratospheric sink for chlorofluoromethanes: chlorine atom-catalysed destruction of ozone, Nature, 249, 810–812, 1974.

Nevas S., Gröbner J., Egli L., and Blumthaler M.: Stray light correction of array spectroradiometers for solar UV measurements, Appl. Opt., 53, 4313–4319, 2014.

Nicolet M.: On the molecular scattering in the terrestrial atmosphere: An empirical formula for its calculation in the homosphere, Planet. Space Sci., 32, 1467–1468, 1984.

Paur R. J. and Bass A. M.: The ultraviolet cross-sections of ozone: II. Results and temperature dependence, Atmospheric Ozone, 611–616, 1985.

Poikonen T., Kärhä P., Manninen P., Manoocheri F., and Ikonen E.: Uncertainty analysis of photometer quality factor f_1', Metrologia, 46, 75–80, 2009.

Reda I. and Andreas A.: Solar Position Algorithm for Solar Radiation Applications, Technical Report NREL/TP-560-34302, Colorado, USA, Revised 2008.

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, 2014.

Redondas A., Rodríguez-Franco J. J., López-Solano J., Carreño V., Leon-Luis S. F., Hernandez-Cruz B., and Berjón A.: The Regional Brewer Calibration Center - Europe: an overview of the X Brewer Intercomparison Campaign, WMO Commission for Instruments and Methods of Observation, 1–8, 2016.

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.: Stratospheric ozone depletion: A review of concepts and history, Reviews of Geophysics, 37, 275–316, 1999.

Toledano C., Cachorro V. E., Berjon A., de Frutos A. M., Sorribas M., de la Morena B. A., and Goloub P.: Aerosol optical depth and Ångström exponent climatology at El Arenosillo AERONET site (Huelva, Spain), Q. J. R. Meteorol. Soc., 133, 795–807, 2007.

UNTC, United Nations Treaty Collection, Chapter XXVII, Environment: *2.a Montreal Protocol on Substances that Deplete the Ozone Layer*, Montreal, 16 September 1987.

Velders G. J. M., Andersen S. O., Daniel J. S., Fahey D. W., and McFarland M.: The importance of the Montreal Protocol in protecting climate, Proc. Natl. Acad. Sci. U.S.A. 104, 4814–4819, 2007.

Weber M., Gorshelev V., and Serdyuchenko A.: Uncertainty budgets of major ozone absorption cross sections used in UV remote sensing applications, Atmos. Meas. Tech., 9, 4459–4470, 2016.

Weber M., Coldewey-Egbers M., Fioletov V. E., Frith S. M., Wild J. D., Burrows J. P., Long C. S., and Loyola D.: Total ozone trends from 1979 to 2016 derived from five merged observational datasets – the emergence into ozone recovery, Atmos. Chem. Phys., 18, 2097–2117, 2018.

Zuber R., Sperfeld P., Nevas S., Sildoja M., and Riechelmann S.: A stray light corrected array spectroradiometer for complex high dynamic range measurements in the UV spectral range, Proc. of the 13th Int. Conf. on New Developments and Applications in Optical Radiometry (NEWRAD 2017), Tokyo, Japan, 13 – 16 June, 2017a.

Zuber R., Sperfeld P., Riechelmann S., Nevas S., Sildoja M., and Seckmeyer G.: Adaption of an array spectroradiometer for total ozone column retrieval using direct solar irradiance measurements in the UV spectral range, Atmos. Meas. Tech., 11, 2477–2484~~(Discussions)~~, 2018~~7b~~.

March 28, 2018

Response letter to Referee #1 on the manuscript "Monte Carlo method for determining uncertainty of total ozone derived from direct solar irradiance spectra: Application to Izaña results"

Authors: Anna Vaskuri, Petri Kärhä, Luca Egli, Julian Gröbner, and Erkki Ikonen

Article reference: amt-2017-403

Authors: We thank Anonymous Referee #1 for the valuable comments that helped us in improving the manuscript. We have included below our detailed responses to all comments.

Specific Comments

"**P1 L14:** "The reason often is that the correlations are not known". It would be better to insert "unknown" into the previous sentence before "correlations" and remove this sentence entirely."

> ***Authors:*** We remove that sentence and move "unknown" to the sentence before. After this change, the sentence is: "One frequently overlooked problem with uncertainty evaluation is that the spectral data may hide systematic wavelength dependent errors due to unknown correlations (Kärhä *et al.* (2017b, 2018); Gardiner *et al.* (1993))."

"**P1 L17:** This would be better phrased as a complete uncertainty budget being necessary to understand long term environmental trends, rather than increased uncertainties improving the reliability of long term trends."

> ***Authors:*** We revise the sentence as: "Complete uncertainty budgets for quantities measured are necessary to understand long term environmental trends, such as changes in the stratospheric ozone concentration (e.g. Molina and Rowland (1974)) and solar UV radiation (e.g. Kerr and McElroy (1993); McKenzie *et al.* (2007))."

"**P2 L5:** Better as "...excluding the remainder of the sky". Also should state field of view of each instrument here or a later point in the field campaign section."

> ***Authors:*** We revise the sentence as: "The field of view of the spectroradiometers has been limited so that they measure direct spectral irradiance of the Sun, excluding most of the indirect radiation from the remainder of the sky."
>
> We included the field of view of each spectroradiometer in Section 2: The field of view with a full opening angle is 2.5° for QASUME (Gröbner *et al.* (2017)), 2.8° for BTS (Zuber *et al.* (2017b)), and 1.5° for AVODOR according to the manual of the collimator tube used, J1004-SMA by CMS Ing.Dr.Schreder GmbH.

"**P2 L6:** This section may be better called "The Izaña field measurement campaign and instrument description", and include some additional details on each instrument - e.g. the field of view and other pertinent details. Alternatively the uncertainty tables would be better moved to later in the manuscript where the individual contributions are discussed."

Authors: We revise the title of Section 2 as: "ATMOZ field measurement campaign and instrument description" and move Tables 1, 2 and 3 to Section 4.

We include more details of the instruments after line 10 on page 3 as:

"The data sets measured by three different spectroradiometers were studied in this work. These spectroradiometers use different techniques to measure the spectral distribution of radiation. Monochromator-based spectroradiometers like QASUME, measure one nearly monochromatic wavelength band at a time, and thus measuring the full spectrum is relatively slow. On the other hand, they usually have significantly better stray light properties than array-based spectroradiometers like BTS and AVODOR, which image the full spectrum at once by dispersing the incoming radiation towards a photodiode array.

QASUME spectroradiometer collects and guides the incoming radiation with input optics and a quartz fiber bundle to the entrance slit of a Bentham DM150 double monochromator (Gröbner *et al.* (2005)). One wavelength at a time can be selected by rotating the two gratings of the double-monochmomator. Then, the monochromatic signal is measured with a photomultiplier tube. QASUME is usually operated in global spectral irradiance mode (Gröbner *et al.* (2005); Hülsen *et al.* (2016)), but during the campaign it was equipped with a collimator tube with a full opening angle of 2.5° allowing the measurement of direct solar spectral irradiance (Gröbner *et al.* (2017)). The measurement range of QASUME during the campaign was limited to 250 nm – 500 nm with a step interval of 0.25 nm, so that one spectrum was measured every 15 minutes. To ensure stable outdoor measurements, the double-monochromator of QASUME was mounted inside a temperature-controlled weather-proofed box (Hülsen *et al.* (2016)).

BTS spectroradiometer utilizes a stationary grating and a back-thinned cooled CCD array detector, mounted in a Czerny-Turner configuration (Zuber *et al.* (2017a, b)). To measure direct solar spectrum, BTS was equipped with a collimator tube with a full opening angle of 2.8° designed by PTB, and it uses an internal filter wheel system with 8 filter positions together with a specific measurement routine to reduce stray light. BTS was mounted on a solar tracker, EKO STR-32G by EKO Instruments Co., Ltd., with pointing accuracy better than 0.01°. A weather-proof housing with temperature control allows BTS operation at the ambient temperatures from –25 °C to 50 °C. During the ATMOZ campaign, the housing temperature of BTS was measured to be stable within 0.1 °C (Zuber *et al.* (2017b)). The measurement range of BTS was 200 nm – 430 nm with a step size of 0.2 nm during the campaign. One spectrum was measured every 45 seconds.

AVODOR spectroradiometer has a stationary grating and a back-thinned cooled CCD array detector in a Czerny-Turner configuration. AVODOR measures the spectrum from 200 nm to 540 nm with a step size of 0.14 nm in the UV region. During the ATMOZ campaign, the field of view of AVODOR was limited to 1.5° by a commercial collimator tube used, J1004-SMA by CMS Ing.Dr.Schreder GmbH. The spectral range of AVODOR was limited between 295 nm and 345 nm

by a combination of two solar blind filters to reduce stray light from the visible and infrared parts of the solar spectrum. The solar blind filters were mounted between the collimator tube and the fiber entrance of the spectroradiometer. One spectrum was measured every 30 seconds."

"**P2 L15:** "mountain Teide ...." » "Mount Teide at *an* altitude..."“

Authors: Corrected according to reviewer's suggestion. The revised sentence reads: "The measurements took place on the Mount Teide at an altitude of 2.36 km above the sea level (28.3090° N, 16.4990° W)."

"**P2 L16** "Station pressure of 772.8 hPa was monitored during the campaign with a standard uncertainty of 1.3 hPa" » "Station pressure was monitored during the campaign and determined to be 772.8 hPa with a standard uncertainty of 1.3 hPa"”

Authors: Corrected according to reviewer's suggestion.

"**P4 L4:** "The tables also give division of the uncertainty components to different correlation types as described in Section 4." » "The tables also attribute uncertainty contributions to different correlation types as described in Section 4." I think this is what you mean, either way needs a rephrasing to clarify."

Authors: We admit this sentence is not clear so we remove it. We move Tables 1, 2 and 3 to Section 4.

"**P10 L2:** "equal *to* unity"”

Authors: Corrected according to reviewer's suggestion. The revised sentence is: "The weights γ_i for the base functions are selected in an *N*-dimensional spherical coordinate system (Hicks and Wheeling (1959)) in such a way that the variance of the final deviation function equals to unity."

"**P10:** This section on the MC description is clearly the core of the study and where the error estimates are derived, but needs more work and clarification. The details and reasoning behind the approach may be in Karha et al 2017, but it would assist reader of this manuscript to relate MC model and, for example, its sinusoidal terms to physical sources of uncertainty, and how these are calculated for random, unfavourable, and full correlations. At present this isn't clear."

Authors: We agree that the core of the paper needs more attention, and we clarify many parts in the text. The text about spectral correlations in the introduction is modified as:

"TOC can be determined from spectral measurements of direct solar UV irradiance (Huber *et al.* (1995)). We have developed a Monte Carlo (MC) based model to estimate the uncertainties of the derived TOC values. One frequently overlooked problem with uncertainty evaluation is that the spectral data may hide systematic wavelength dependent errors due to unknown correlations (Kärhä *et al.* (2017b, 2018); Gardiner *et al.* (1993)). Omitting possible correlations may lead into underestimated uncertainties for derived quantities, since spectrally varying systematic errors typically produce larger deviations than uncorrelated noise-like variations that traditional uncertainty estimations predict. Complete uncertainty budgets for quantities measured are necessary to understand long term environmental trends, such as changes in the stratospheric

ozone concentration (e.g. Molina and Rowland (1974)) and solar UV radiation (e.g. Kerr and McElroy (1993); McKenzie *et al.* (2007)).

Physically, correlations may originate, e.g., from lamps or other light sources used in calibrations. If their temperatures change e.g. due to ageing or current setting, a spectral change in the form of Planck's radiation law is introduced. Non-linearity in the responsivity of a detector causes systematic differences between high and low measured values. The introduced spectrally systematic but unknown changes in irradiance may change the derived TOC values significantly, exceeding the uncertainties calculated assuming that the uncertainty in irradiance behaves like noise. The presence of correlations in measurements can be seen in many ways. For example, problems have occurred when new ozone absorption cross-sections have been taken into use (Redondas *et al.* (2014); Fragkos *et al.* (2015)). Derived ozone values may change significantly because different systematic errors are included in the different cross-sections. Also, TOC estimated from a measured spectrum often depends on the wavelength region chosen, although the measurement region should not affect the result much."

Regarding the above paragraphs, a new reference was included in the AMT Discussion manuscript:

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, 2014.

Near Tables 1 – 3 about the uncertainty budgets of the three spectroradiometers, which were moved to Section 4, we place text about how the uncertainties and correlations have been estimated:

"The uncertainties due to radiometric calibration include factors such as the uncertainty of the standard lamp used, and the additional uncertainty due to noise and alignment. QASUME has been validated using various methods, thus the uncertainty due to calibration is low (Hülsen *et al.* (2016)). For QASUME and BTS, we assume the correlations to be equally distributed between *full* correlation, *unfavourable* correlation, and *random* correlation (Kärhä *et al.* (2018)). Spectra measured with AVODOR are significantly noisier, thus half of the uncertainty is associated to the *random* component. Values for instability of the calibration lamp are based on long-term monitoring. The lamp irradiances have been noted to gradually drop with no significant wavelength dependent structure within the wavelength region concerned. Non-linearity values are estimations of the operators of the devices. Non-linearity is typically manifested so that the responsivity of the device changes gradually from high readings to low readings. This can cause significant change in the TOC values, thus we assume the correlation type to be *unfavourable*. Uncertainties due to device stability and temperature dependence are based on long-term monitoring. The changes have been found to be independent on wavelength in the region concerned, thus *full* correlation is assumed. Noise is the average standard deviation of typical measurements at noon over the wavelength region concerned. The wavelength scales of the devices have been checked using emission lines of gas discharge lamps. The uncertainty values

given are the estimated standard deviations of the possible remaining errors after corrections. Wavelength error can introduce a significant change in TOC, because it introduces an error in the form of the derivative of the spectral irradiance. Thus, *unfavourable* correlation is assumed. Most of the uncertainty components are slightly wavelength dependent but to simplify simulations, average uncertainty values are used over the wavelength range between 300 nm and 340 nm."

Regarding the paragraph above, one reference was updated in the AMT Discussion manuscript:

Kärhä P., Vaskuri A., Pulli T., and Ikonen E.: Key comparison CCPR-K1.a as an interlaboratory comparison of correlated color temperature, J. Phys.: Conf. Ser., 972, 012012, 2018. doi:10.1088/1742-6596/972/1/012012

To clarify the uncertainty components in Table 4, we also include new sentences in the AMT Discussion manuscript before line 5 on page 14:

"For components (a) – (d) in Table 4, the mechanism of contributing to the uncertainty of TOC is known. We know the standard uncertainty of the O_3 layer altitude of 26 km to be u = 0.5 km, so we vary the altitude accordingly and note the variance of the resulting TOC."

"**P16 L3:** If Brewer #183 is included as a reference instrument, then it would be useful to include a similar uncertainty budget for this instrument, even if only summarised. Also the community would find it useful to put these results into context and comparison with those observed at instrument intercomparisons, and often quoted as a measure of instrument or data quality. i.e. for Brewers is the actual uncertainty determined by the MC methodology much large than expected from the intercomparison error, and, what is the primary source (so efforts can be made to reduce it)."

Authors: We thank the referee for the suggestion to include an overall uncertainty budget of Brewer #183. However, this study focuses on MC uncertainty calculation from full spectrum ozone retrieval. An overall uncertainty budget of the double ratio technique ozone retrieval (e.g. for Brewers and Dobsons) is subject of another publication from the ATMOZ field measurement campaign, which is under preparation, but not citable yet. The Brewer data shown in this publication serve only for comparison of the retrieved ozone during the campaign.

Additional notes by the authors

As we modified the retrieval algorithm by including a new offset factor c to compensate for full spectral deviations, the results compared with the AMT Discussion paper will slightly change. Please, see separate response letter addressed to Referee #2 for full details.

March 22, 2018

Response letter to Referee #2 on the manuscript "Monte Carlo method for determining uncertainty of total ozone derived from direct solar irradiance spectra: Application to Izaña results"

Authors: Anna Vaskuri, Petri Kärhä, Luca Egli, Julian Gröbner, and Erkki Ikonen

Article reference: amt-2017-403

Authors: We thank Anonymous Referee #2 for carefully reading our AMT Discussion paper and giving constructive criticism. We will later answer point by point all details raised, but at this stage we clarify the scientific concerns raised.

Comments on Introduction

"First, the paper was probably written very quickly and important information is missing, which makes the understanding quite difficult for the unexperienced reader:

- a complete introduction about the importance of ozone measurements and the assessment of their uncertainty should be included (it is only the first line, so far);

- it is not explained why spectral data should contain correlations (physical basis);"

> Authors: We revised Introduction and included new paragraphs as the referee suggested. After these amendments, the introduction is:
>
> "Atmospheric ozone has been defined as an essential climate variable in the global climate observing system (GCOS-200 (2016)) of the World Meteorological Organization (WMO). Its long-term monitoring is necessary to document the expected recovery of the ozone layer due to the implementation of the Montreal protocol (UNTC (1987)) and its amendments on the protection of the ozone layer. Atmospheric ozone, first discovered by Fabry and Buisson (1913), protects the humans, the biosphere, and infrastructures from adverse effects of ultraviolet (UV) radiation by shielding the Earth surface from excessive radiation levels (McElroy and Fogal (2008)). Since the 1970's, it is known that human-produced chlorofluorocarbons (CFCs) destruct atmospheric ozone (Molina and Rowland (1974)) and have led to recurring massive losses of total ozone in the Antarctic in the form of the ozone hole (Farman et al. (1985); Solomon et al. (1986)). An unprecedented ozone depletion has also been recently observed in the Arctic (Manney et al. (2011)), while in the middle-latitudes, moderate ozone depletion has been observed (Solomon (1999)). The Montreal protocol and its amendments have been successful in reducing the emission of ozone-depleting substances (Velders et al. (2007)). Nevertheless, recent studies give conflicting results with respect to the observation of a recovery of the ozone layer, and model projections have shown the recovery to occur not before the middle of the 21st century (Ball et al. (2018); Weber et al. (2018)). Therefore, careful monitoring of the thickness of the ozone layer with uncertainties of 1% or less is crucial in verifying the successful implementation of the Montreal Protocol and the eventual recovery of the ozone layer to pre-1970's levels.

"Traceability for atmospheric total column ozone (ATMOZ)" was a three-year project funded jointly by the European Metrology Research Programme (EMRP) and the European Union (ATMOZ project (2014 – 2017)). The goal of this project was to produce traceable measurements of total ozone column (TOC) with uncertainties down to 1%, by a systematic investigation of the radiometric and spectroscopic aspects of the methodologies used in retrieval. Another objective of the project was to provide a comprehensive treatment of uncertainties of all parameters affecting the TOC retrievals using spectrophotometers. This paper presents outcome of the work on studying the uncertainty of TOC obtained from spectral direct solar irradiance measurements, taking correlations among wavelengths explicitly into account.

TOC can be determined from spectral measurements of direct solar UV irradiance (Huber et al. (1995)). We have developed a Monte Carlo (MC) based model to estimate the uncertainties of the derived TOC values. One frequently overlooked problem with uncertainty evaluation is that the spectral data may hide systematic wavelength dependent errors due to unknown correlations (Kärhä et al. (2017b, 2018); Gardiner et al. (1993)). Omitting possible correlations may lead into underestimated uncertainties for derived quantities, since spectrally varying systematic errors typically produce larger deviations than uncorrelated noise-like variations that traditional uncertainty estimations predict. Complete uncertainty budgets for quantities measured are necessary to understand long term environmental trends, such as changes in the stratospheric ozone concentration (e.g. Molina and Rowland (1974)) and solar UV radiation (e.g. Kerr and McElroy (1993); McKenzie et al. (2007)).

Physically, the correlations may originate, e.g., from lamps or other light sources used in calibrations. If their temperatures change e.g. due to ageing or current setting, a spectral change in the form of Planck's radiation law is introduced. Non-linearity in the responsivity of detector causes systematic differences between high and low measured values. The effect of the introduced spectrally systematic, but unknown changes on the TOC values derived from the affected spectra may deviate significantly from the uncertainty calculated assuming that the error behaves like noise. The presence of correlations in measurements can be seen in many ways. For example, problems have occurred when new ozone absorption cross-sections have been taken into use (Redondas et al. (2014); Fragkos et al. (2015)). Derived ozone values may change significantly because different systematic errors are included in the different cross-sections. Also, TOC estimated from a measured spectrum often depends on the wavelength region chosen, although the measurement region should not affect the result much.

In this paper, we introduce a new method for dealing with possible correlations in spectral irradiance data and analyse uncertainties in ozone retrievals for three different spectroradiometers used in a recent intercomparison campaign at Izaña, Tenerife, to demonstrate how it can be used in practice. One of the instruments is QASUME (Gröbner et al. (2005)) that is the World reference UV spectroradiometer at the World Radiation Center (PMOD/WRC). The second one is an array-based high-quality spectroradiometer BTS2048-30-UV-S-WP (BTS) from Gigahertz-Optik (Zuber et al. (2017a, b)), operated by PTB. The third one is an array-based spectroradiometer AvaSpec-ULS2048LTEC (AVODOR) from Avantes, operated by PMOD/WRC. The field of view of the spectroradiometers has been limited so that they measure direct spectral irradiance of the Sun, excluding most of the indirect radiation from the remainder of the sky. "

New references were included:

Ball W. T., Alsing J., Mortlock D. J., Staehelin J., Haigh J. 5 D., Peter T., Tummon F., Stübi R., Stenke A., Anderson J., Bourassa A., Davis S. M., Degenstein D., Frith S., Froidevaux L., Roth C., Sofieva V., Wang R., Wild J., Yu P., Ziemke J. R., and Rozanov E. V.: Evidence for a continuous decline in lower stratospheric ozone offsetting ozone layer recovery, Atmos. Chem. Phys., 18, 1379–1394, 2018.

Fabry C. and Buisson H.: L'absorption de l'ultra-violet par l'ozone et la limite du spectre solaire, J. de Phys., 3, 196–206 (1913).

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

GCOS-200: The Global Observing System for Climate: Implementation Needs (World Meteorological Organization, 2016).

Manney G. L., Santee M. L., Rex M., Livesey N. J., Pitts M. C., Veefkind P., Nash E. R., Wohltmann I., Lehmann R., Froidevaux L., Poole L. R., Schoeberl M. R., Haffner D. P., Davies J., Dorokhov V., Gernandt H., Johnson B., Kivi R., Kyrö E., Larsen N., Levelt P. F., Makshtas A., McElroy C. T., Nakajima H., Concepción Parrondo M., Tarasick D. W., von der Gathen P., Walker K. A., and Zinoviev N. S.: Unprecedented Arctic ozone loss in 2011, Nature, 478, 469–475, 2011.

McElroy C.T. and Fogal P.F., Ozone: From Discovery to Protection, Atmosphere-Ocean, 46, 1–13, 2008.

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, 2014.

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.: Stratospheric ozone depletion: A review of concepts and history, Reviews of Geophysics, 37, 275–316, 1999.

UNTC, United Nations Treaty Collection, Chapter XXVII, Environment: 2.a Montreal Protocol on Substances that Deplete the Ozone Layer, Montreal, 16 September 1987.

Velders G. J. M., Andersen S. O., Daniel J. S., Fahey D. W., and McFarland M.: The importance of the Montreal Protocol in protecting climate, Proc. Natl. Acad. Sci. U.S.A. 104, 4814–4819, 2007.

Weber M., Coldewey-Egbers M., Fioletov V. E., Frith S. M., Wild J. D., Burrows J. P., Long C. S., and Loyola D.: Total ozone trends from 1979 to 2016 derived from five merged observational datasets – the emergence into ozone recovery, Atmos. Chem. Phys., 18, 2097–2117, 2018.

SPECIFIC COMMENT 1 - Retrieval model

"It should be explained why the retrieval method (Eq. 8) was chosen. An obvious drawback is that this method is not invariant for spectrally-constant factors ("full correlation" or systematic errors in the absolute calibration, as hypothesised for AVODOR). It should be stated what networks / instruments use this model. For example, it is false that "This approach is consistent with the ozone measurements with Brewer" (p. 7, line 17): the Brewer algorithm is invariant for "full spectral correlation" (therefore, the "full correlation" term would not make sense with a different retrieval method, e.g. for a method where the offset is also included in a DOAS-like fit)."

Authors: Omitting an offset factor in our model was a mistake from our side, as we did not take into account how easily full correlations appear in solar UV irradiance measurements. As the results of AVODOR show, it is clearly needed. We thus modified the atmospheric model as

$$E_{\mathrm{s}}(\lambda) = c \cdot E_{\mathrm{ext}}(\lambda) \cdot \exp\left[-\alpha_{\mathrm{O_3}}(\lambda, T_{\mathrm{eff}}) \cdot \mathrm{TOC} \cdot m_{\mathrm{TOC}} - \tau_{\mathrm{R}}(\lambda, P_0, z_0, \phi) \cdot m_{\mathrm{R}} - \tau_{\mathrm{AOD}}(\lambda) \cdot m_{\mathrm{AOD}}\right], \quad (1\mathrm{R})$$

$$\tau_{\mathrm{AOD}}(\lambda) = \beta \cdot \lambda^{-\alpha}, \quad (2\mathrm{R})$$

where the parameters are as in Eqs. (3)–(7) in the AMT Discussion manuscript and new multiplier c is the scaling factor set as an additional free parameter. Thus, there are three free parameters; TOC, $\beta \geq 0$, and c to be fitted. We set $\beta \geq 0$ since to our understanding aerosols attenuate the direct solar UV spectrum. The offset factor will correct for any wavelength independent deviation in the measurements. The results of the paper change due to this modification. The large offset of AVODOR will diminish. The other devices will also have small changes in their results. We give below discussion of the changed results.

We now call $E_{\mathrm{s}}(\lambda)$ as the modelled direct spectral irradiance at the Earth surface, and β as Ångström turbidity coefficient, although Ångström (1964) himself called β as extinction coefficient.

We also studied the effect of using a linear approximation of aerosol optical depth in TOC values estimated as it was used e.g. by Huber et al. (1995). In this test, Eq. (2R) was replaced with

$$\tau_{\mathrm{AOD}}(\lambda) = a \cdot (\lambda - 340\ \mathrm{nm}) + b, \quad (3\mathrm{R})$$

where a and b are free parameters without bound constraints. When linear AOD model in Eq. (3R) is used with (1R), the multiplier needs to be removed by setting it to $c = 1$, because the free parameter b produces almost similar multiplier ($e^{-b \cdot m_{\mathrm{AOD}}}$). Based on our analysis, the linear AOD model gave almost similar results as Eq. (2R), so we prefer the Ångström AOD model and an offset factor c, as we believe this approach is more physical.

"Also, since this method gives more weight to lower irradiances, the authors should carefully explain how the "wavelength region where the signal is above the noise floor" was determined (5E-3 and 1E-5 noise floors);"

Authors: It is not obvious which least squares minimization method works best, as the method of obtaining TOC from a measured spectrum is very critical to the way that the least squares fitting is performed. We have now studied different options how to perform the weighting. To facilitate discussion, we rewrite Eq. (8) in our AMT Discussion manuscript as

$$S = \sum_{i=1}^{n} w(\lambda_i)[E_{\mathrm{s}}(\lambda_i) - E(\lambda_i)]^2, \quad (4\mathrm{R})$$

where $w(\lambda_i)$ is the weight.

In our AMT Discussion manuscript, we used relative errors in calculation, i.e., $w(\lambda_i) = E(\lambda_i)^{-2}$. This method was also used in the article by Huber et al. (1995). This method works very well with high-accuracy double-monochromator instruments giving several orders of magnitude of useful data down to 10^{-6} $Wm^{-2}nm^{-1}$ at UV region, such as the one used in the original study by Huber et al. (1995), or the QASUME used in our experiment. However, with large zenith angles in the morning and in the evening, the spectra have to be cut at some reasonable threshold level even with monochromator-based instruments when the relative least squares minimization is used.

With array spectroradiometers that suffer from stray light and high baseline noise as presented in Fig. 1R, the dynamic range can easily be less than two orders of magnitude. Figure 1R shows the fitted spectra for all three devices, and the residuals of the fits. The residuals show clearly how the stray light is present in both BTS and AVODOR. The threshold level where the data is cut affects the TOC results at large zenith angles significantly, when the relative least squares minimization is used as demonstrated in Figure 2R.

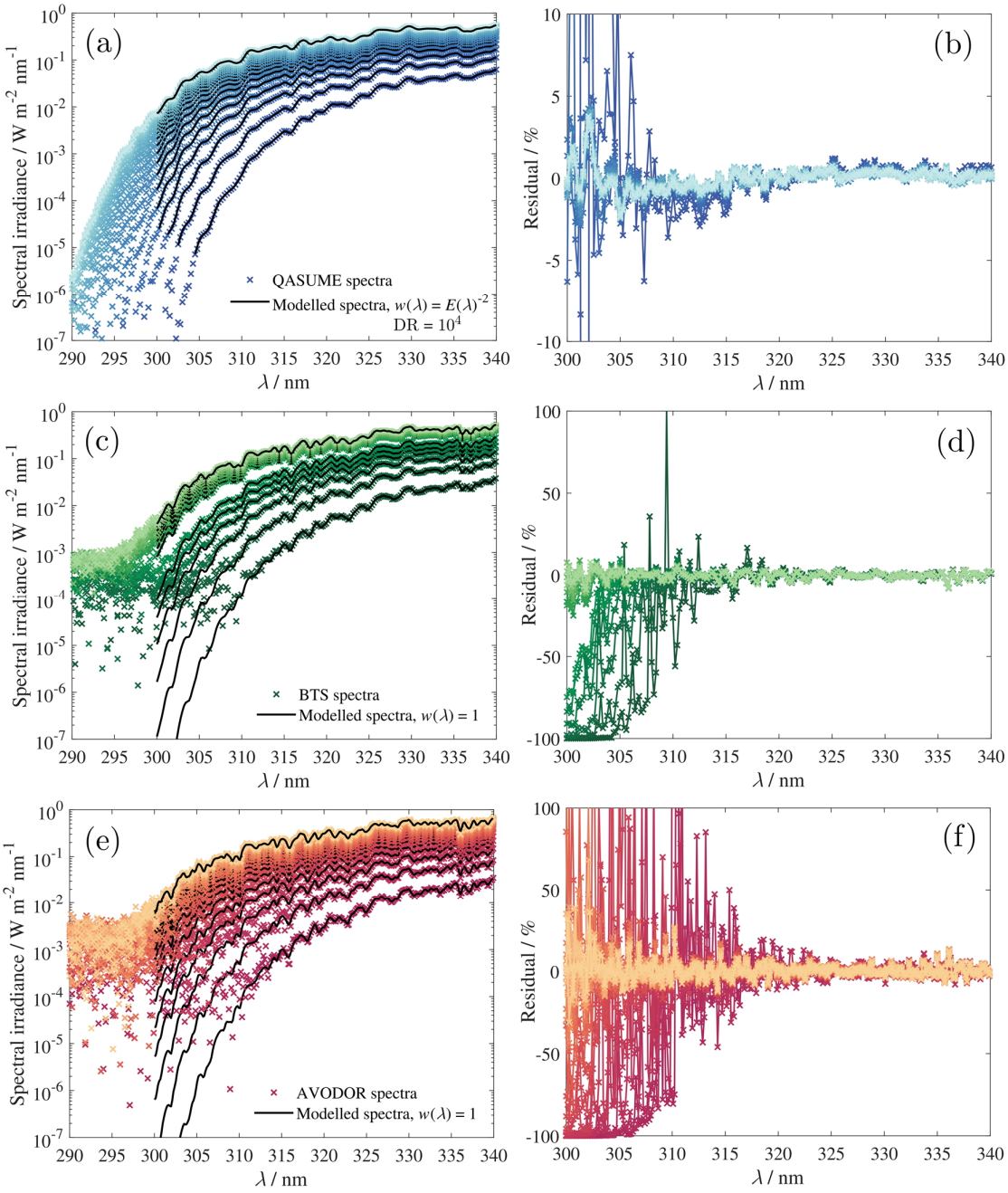


Figure 1R. Examples of fitting the atmospheric model to the direct ground-based solar UV spectra between 300 nm and 340 nm for QASUME (a–b), BTS (c–d) and AVODOR (e–f). In figures on the left hand side, the coloured symbols indicate measured spectra, and the black solid curves indicate modelled spectra. Figures on the right hand side show the relative spectral residuals of the fits. In (a), the abbreviation DR refers to the dynamic range of QASUME data used in the least squares fitting.

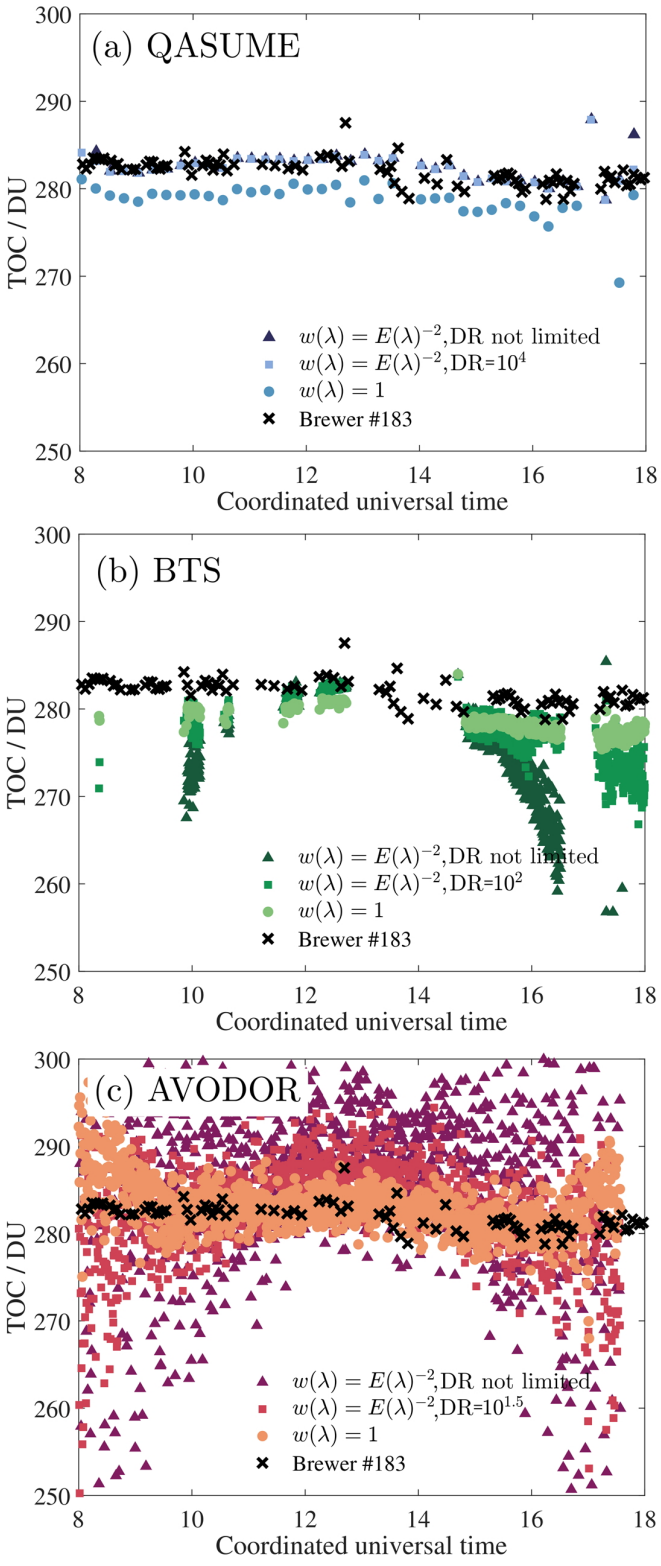


Figure 2R. TOC values estimated using different weighting in least squares minimisation and using Ångström AOD model for QASUME (a), BTS (b) and AVODOR (c). TOC values for Brewer #183 are plotted as black crosses for comparison. The colour codes and the associated figure legends denote the weighting used (Eq. (4R)) and the dynamic range DR used.

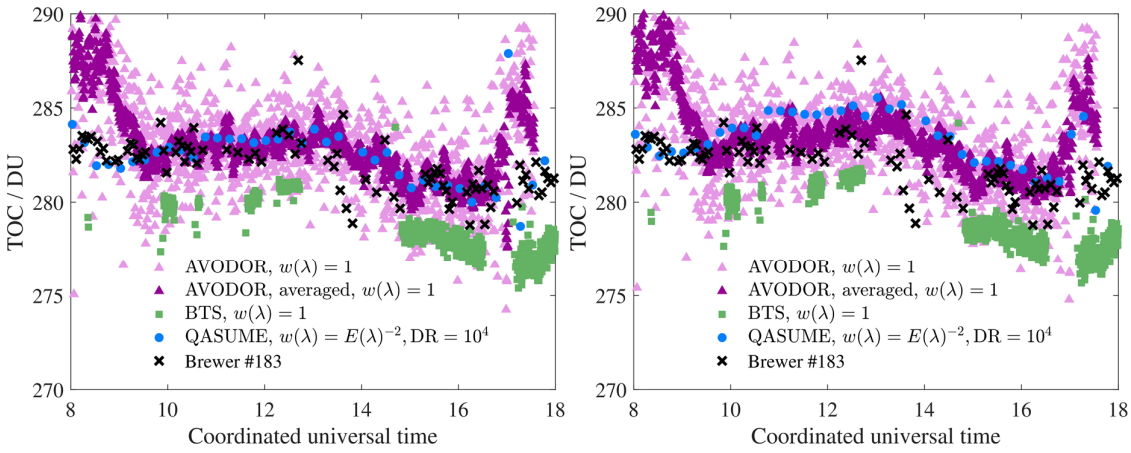
Figure 2R shows TOC values calculated from spectra throughout the day using various minimization methods. As can be seen, using relative least squares fitting causes significant inverse U-shape to the results obtained with BTS and AVODOR, which is not present in the QASUME data (triangle and square symbols). In the case of BTS and AVODOR, the TOC results change significantly by changing the threshold level. This can be seen in the curves, where the dynamic range of the data was 10^2 for BTS and $10^{1.5}$ for AVODOR. However, the stray light still causes U-shape to the curves. In our opinion, the most objective method for analysing the results of the array based spectroradiometers is to use absolute least squares minimisation by setting the weight to $w(\lambda) = 1$ (light green and orange circles in Fig. 2R). The absolute least squares minimization does not give much weight to the baseline noise tail. As can be seen in Fig. 2R, with the absolute least squares fitting there is very little variation throughout the day and we do not have to subjectively limit the dynamic range the spectra.

As we know from the results of Brewer and QASUME e.g. in Fig. 2R, the TOC was rather constant during the measurement day. Thus, the inverse U-shape notable for most curves is a property of the analysis method originating from the stray light and the baseline noise that our atmospheric model cannot account for. Herman et al. (2015) have also reported the inverse U-shape in TOC due to stray light of array spectroradiometers.

Our conclusion is that the relative least squares minimization (Eq. (4R) with $w(\lambda) = E(\lambda)^{-2}$) is only applicable to spectroradiometers free from stray light. For our paper, we choose to use this approach to QASUME with the dynamic range of four orders of magnitude. With high-accuracy instruments this is justified, because such weighting uses all data available. For both array spectroradiometers, BTS and AVODOR, we use absolute least squares minimization (Eq. (4R) with $w(\lambda) = 1$). This approach is justified, as it gives less weight to the low irradiance values distorted by stray light.

According to Fig. 2R the array spectroradiometers operate well at noon, even if analysed using relative least squares fitting. Comparing results at noon shows that using absolute least squares fitting underestimates the TOC values by 2 DU as compared to relative least squares fitting. The same offset, independent of zenith angle can be seen with QASUME.

Figure 3R shows the final comparison results. As can be seen, the inverse U-shape has diminished in both BTS and AVODOR data as compared to Fig. 5 of our AMT Discussion paper, and the agreement is rather good for all instruments.



(a) TOC estimated using Eq. (1R) with Ångström AOD model in Eq. (2R)

(b) TOC estimated using Eq. (1R) with linear AOD model in Eq. (3R)

Figure 3R. Absolute TOC values estimated for the spectroradiometers studied. Average of 10 neighbouring values has been included for AVODOR to show the spectral shape behind the noise.

In our revised manuscript, we will include the presented Figures 1R – 3R, and their descriptions. Relative least squares fitting is used with QASUME using the dynamic threshold of four orders of magnitude which can be justified by the extremely low noise level as seen in Figure 1R. For the array spectrometers, we use the absolute least squares fitting. In addition, to avoid confusion we removed the sentence stating that our approach is consistent with Brewer algorithm.

A new reference was included in the manuscript:

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.

"For example, it is false that "This approach is consistent with the ozone measurements with Brewer" (p. 7, line 17): the Brewer algorithm is invariant for "full spectral correlation" (therefore, the "full correlation" term would not make sense with a different retrieval method, e.g. for a method where the offset is also included in a DOAS-like fit)."

It is true that with our revised MC uncertainty method, "full correlation" does not contribute, to the uncertainty of TOC. Full correlation in the ozone absorption cross section is an exception to this, since variable c cannot completely compensate for errors in the exponent. However, it is still meaningful to divide the components to full, unfavourable and random correlations. The part that is fully correlated does not contribute to uncertainty, thus the resulting uncertainty will be smaller correspondingly. For example, in Table 4 of our AMT Discussion manuscript, the spectral uncertainty arising from full spectral correlation of the extra-terrestrial spectrum does not contribute to the uncertainty in TOC, but it contributes to the total uncertainty of the extra-terrestrial spectrum.

SPECIFIC COMMENT 2 - Uncertainty model

"The uncertainty model (Eq. 9-10) is very complex. However, for the most part of the paper, only three terms of the Fourier series are used. Thus, I wonder why such a complex initial framework must be described."

> Authors: The Fourier series approach is intended as a generic tool that can be applied to any quantity derived mathematically from a spectrum. This is explained on lines 19–20 on page 10 of the AMT Discussion manuscript. In the publication by Kärhä et al. (2018), it is demonstrated that many national metrology institutes have systematic wavelength dependent deviations in their spectral irradiance scales. The Fourier series can reproduce such spectral shapes. The main benefit of using the Fourier series is that the variance of the systematic spectral deviation can easily be controlled, as introduced in Section 4.1 of our AMT Discussion manuscript.
>
> An updated reference:
>
> Kärhä P., Vaskuri A., Pulli T., and Ikonen E.: Key comparison CCPR-K1.a as an interlaboratory comparison of correlated color temperature, J. Phys.: Conf. Ser. 972, 012012, 2018. doi:10.1088/1742-6596/972/1/012012

"Also, it should be explained why deviations in measurements should follow this model. Coming to the "unfavourable correlation", "The obtained TOC value is affected most by spectral distortion that mimics the spectral shape of the ozone absorption... The first combination of constant offset and one sinusoidal function ... is closest to this extreme"."

> Authors: As it is explained on lines 19–20 on page 10 of the AMT Discussion manuscript, the model does not assume any particular spectral shape of the deviation. Fourier series has a property that in its full form it can mimic any shape of deviation. This takes place at the Nyquist criterion, where N is equal to half the number of wavelengths available. Also with smaller values of N, the MC parameters can account for unknown spectral shapes of lower complexity.
>
> We sum the base functions starting with the simplest forms, N = 1, 2 … The logic here is that when one studies which kind of correlations there are present in typical measurements, correlations such as these appear commonly. This can be seen in (Kärhä et al. (2018)) and (Woolliams et al. (2006)) which examine deviations in the spectral irradiance measurements of national standards laboratories.
>
> A reference:
>
> Woolliams E. R., Fox N. P., Cox M. G., Harris P. M. and Harrison N. J.: Final report on CCPR-K1-a: Spectral irradiance from 250 nm to 2500 nm, Metrologia, 43, Technical Supplement 02003, 2006.

"However, from Eq. 10, this term varies with lambda1 and lambda2, so the width of the sinusoid is different for the three instruments and comparison of the results is difficult."

> Authors: This is true. We now changed our analysis so that all instruments use the same wavelength region 300 nm – 340 nm and remodelled the results.

"Also, why should this term be a full sinusoid cycle? Why should the period of the oscillation be the same for all uncertainty components (calibration, measurement, cross sections, etc.)?"

Authors: As it is explained on lines 22–25 on page 18 of the AMT Discussion manuscript, simpler forms of deviations could be used after the general solution of the problem is known.

In our method, adding half a sinusoid or a slope as a base function in the general solution would lead into more complicated calculations when determining the rest of the base functions, as all the base functions need to be orthogonal with respect to each other. It is possible to use other orthogonal sets of functions instead of sinusoidal functions, but that would involve more complicated mathematics.

"Finally, it is obvious that phi has an enormous role for the "unfavourable correlation" term: depending on phi's value (i.e. similarity with ozone spectral cross section), the induced error could be huge or negligible. The role of phi should be explained better. Physically, what does phi represent? It is expected to randomly change in one instrument or not?"

Authors: Parameter ϕ_i is used as an equally distributed MC variable between 0 and 2π (see page 9 lines 29–30 of the AMT Discussion manuscript). It is needed in the Fourier series so that it can reproduce all possible spectral shapes.

March 30, 2018

Second response letter to Referee #2 on the manuscript "Monte Carlo method for determining uncertainty of total ozone derived from direct solar irradiance spectra: Application to Izaña results"

Authors: Anna Vaskuri, Petri Kärhä, Luca Egli, Julian Gröbner, and Erkki Ikonen

Article reference: amt-2017-403

Authors: We thank Anonymous Referee #2 for the valuable comments that helped us in improving the manuscript. We have included below our detailed responses to the remaining comments. Our response to the comments regarding the introduction and the specific comments 1 and 2 can be found in our first response letter addressed to Referee #2.

COMMENTS

"the only characteristics of the instruments described in the manuscript are their spectral range. I believe that a study of the instrumental uncertainties should provide a thorough description of the instruments"

> ***Authors:*** To provide more information on the campaign and better description of the instruments, we rename Section 2 as "ATMOZ field measurement campaign and instrument description". The following paragraphs after line 10 on page 3 now describe the instruments used:
>
> "The data sets measured by three different spectroradiometers were studied in this work. These spectroradiometers use different techniques to measure the spectral distribution of radiation. Monochromator-based spectroradiometers like QASUME, measure one nearly monochromatic wavelength band at a time, and thus measuring the full spectrum is relatively slow. On the other hand, they usually have significantly better stray light properties than array-based spectroradiometers like BTS and AVODOR, which image the full spectrum at once by dispersing the incoming radiation towards a photodiode array.
>
> QASUME spectroradiometer collects and guides the incoming radiation with input optics and a quartz fiber bundle to the entrance slit of a Bentham DM150 double monochromator (Gröbner *et al.* (2005)). One wavelength at a time is selected by rotating the two gratings of the double-monochmomator. Then, the monochromatic signal is measured with a photomultiplier tube. QASUME is usually operated in global spectral irradiance mode (Gröbner *et al.* (2005); Hülsen *et al.* (2016)), but during the campaign it was equipped with a collimator tube with a full opening angle of 2.5° allowing the measurement of direct solar spectral irradiance (Gröbner *et al.* (2017)). The measurement range of QASUME during the campaign was limited to 250 nm – 500 nm with a step interval of 0.25 nm, so that one spectrum was measured every 15 minutes. To ensure stable outdoor measurements, the double-monochromator of QASUME was mounted inside a temperature-controlled weather-proofed box (Hülsen *et al.* (2016)).

BTS spectroradiometer utilizes a stationary grating and a back-thinned cooled CCD array detector, mounted in a Czerny-Turner configuration (Zuber *et al.* (2017a, b)). To measure direct solar spectrum, BTS was equipped with a collimator tube with a full opening angle of 2.8° designed by PTB, and it uses an internal filter wheel system with 8 filter positions together with a specific measurement routine to reduce stray light. BTS was mounted on a solar tracker, EKO STR-32G by EKO Instruments Co., Ltd., with pointing accuracy better than 0.01°. A weather-proof housing with temperature control allows BTS operation at the ambient temperatures from –25 °C to 50 °C. During the ATMOZ campaign, the housing temperature of BTS was measured to be stable within 0.1 °C (Zuber *et al.* (2017b)). The measurement range of BTS was 200 nm – 430 nm with a step size of 0.2 nm during the campaign. One spectrum was measured every 45 seconds.

AVODOR spectroradiometer has a stationary grating and a back-thinned cooled CCD array detector in a Czerny-Turner configuration. AVODOR measures the spectrum from 200 nm to 540 nm with a step size of 0.14 nm in the UV region. During the ATMOZ campaign, the field of view of AVODOR was limited to 1.5° by a commercial collimator tube used, J1004-SMA by CMS Ing.Dr.Schreder GmbH. The spectral range of AVODOR was limited between 295 nm and 345 nm by a combination of two solar blind filters to reduce stray light from the visible and infrared parts of the solar spectrum. The solar blind filters were mounted between the collimator tube and the fiber entrance of the spectroradiometer. One spectrum was measured every 30 seconds."

"a Monte carlo model was employed, but important details such as the number of samples that were used, the obtained statistical distribution, etc. are not mentioned;"

Authors: We add new sentences after line 9 on page 10:

"Each standard uncertainty of TOC in Fig. 4 was estimated from the MC results obtained by running the TOC retrieval 1000 times so that the phases ϕ_i and the weights γ_i of the base functions were independent at each round. Retrieved TOC deviations resemble Gaussian distribution when the order of complexity of the deviation function is $N \geq 2$."

"the uncertainty components written in the tables are not properly motivated in Sect. 5. Each number should be accompanied by a clear explanation;"

Authors: We amend the text describing Tables 1, 2, and 3 to describe all components included. The tables will be located in chapter 4 in the new manuscript:

"The uncertainties due to radiometric calibration include factors such as the uncertainty of the standard lamp used, and the additional uncertainty due to noise and alignment. QASUME has been validated using various methods, thus the uncertainty due to calibration is low (Hülsen *et al.* (2016)). For QASUME and BTS, we assume the correlations to be equally distributed between *full* correlation, *unfavourable* correlation, and *random* correlation (Kärhä *et al.* (2018)). Spectra measured with AVODOR are significantly noisier, thus half of the uncertainty is associated to the *random* component. Values for instability of the calibration lamp are based on long-term monitoring. The lamp irradiances have been noted to gradually drop with no significant wavelength structure within the wavelength region concerned. Non-linearity values are estimations of the operators of the devices. Non-linearity is typically manifested so that the responsivity of the device changes gradually from high readings to low readings. This can cause significant change in the TOC values, thus we assume the correlation type to be

unfavourable. Uncertainties due to device stability and temperature dependence are based on long-term monitoring. The changes have been found to be independent on wavelength in the region concerned, thus *full* correlation is assumed. Noise is the average standard deviation of typical measurements at noon over the wavelength region concerned. The wavelength scales of the devices have been checked using emission lines of gas discharge lamps. The uncertainty values given are the estimated standard deviations of the possible remaining errors after corrections. Wavelength error can introduce a significant change in TOC, because it introduces an error in the form of the derivative of the spectral irradiance. Thus, *unfavourable* correlation is assumed. Most of the uncertainty components are slightly wavelength dependent but to simplify simulations, average uncertainty values are used over the wavelength range between 300 nm and 340 nm."

"the discussion about the deviation of the three instruments is very superficial (e.g., "... the reason a the systematic deviation either in the linearity, stray light properties, or the calibration of the device") and inconclusive. How is it connected to the main topic of the paper?"

Authors: Thanks to the improved atmospheric model and the improved TOC results, we can remove that sentence. Please, see the first response letter addressed to Referee #2 for full details.

"language inaccuracies are listed in the "Technical corrections" section. One for all: "Izaña results" in the title is very generic. At the Izaña atmospheric observatory (not simply "Izaña"), several activities are organised and the title should appropriately tell which campaign was taken into consideration;"

Authors: We change the title of our AMT Discussion manuscript as "Monte Carlo method for determining uncertainty of total ozone derived from direct solar irradiance spectra: The results of ATMOZ field measurement campaign at the Izaña Atmospheric Observatory".

"there is a persistent interchange of terms that should not be mixed: uncertainty, deviation, error, etc. (e.g., "uncertainty induced by deviation", p. 10 line 9);"

Authors: In our opinion, we have used those terms correctly, except for "error function" that we change to "deviation function" so that it cannot be confused with "Gauss error function" that is often called simply "error function" or "erf".

Uncertainty is an estimate of the possible error that there may be in a quantity. Spectral errors may further contain systematic deviations. In our analysis we go through systematic deviations (errors) that the uncertainties of the input quantities permit. The variances of TOC obtained with the deviated spectra then give uncertainties for the output quantity TOC. Description of the analysis requires using terms:

- **Standard uncertainty** is the square root of the variance of a probability distribution.
- **Expanded uncertainty** specifies the value of the measurand with 95% confidence. For a Gaussian probability distribution, expanded uncertainty is twice the standard uncertainty ($k = 2$).
- **Error** is a discrepancy between the measured value and the "true" value.
- Due to the nature of Monte Carlo method, in each Monte Carlo round, we introduce an arbitrary **spectral deviation** in the input parameter and observe how it changes TOC value. When we run Monte Carlo TOC retrieval multiple times, we can calculate the standard uncertainty from the variation in the output TOC.

"QASUME is not only a "high-quality reference instrument ... at PMOD/WRC": it is the World reference UV spectroradiometer! Anyway, it should be explained how a global irradiance instrument could measure direct solar irradiance spectra (p. 2 line 5: how was the field of view "limited"?);"

Authors: We revise the sentence on page 2, line 1 as:

"One of the instruments is QASUME (Gröbner *et al.* (2005)) that is the World reference UV spectroradiometer at the World Radiation Center (PMOD/WRC)."

We also include a new paragraph about QASUME in Section 2:

"QASUME spectroradiometer collects and guides the incoming radiation with input optics and a quartz fiber bundle to the entrance slit of a Bentham DM150 double monochromator (Gröbner *et al.* (2005)). One wavelength at a time is selected by rotating the two gratings of the double-monochmomator. Then, the monochromatic signal is measured with a photomultiplier tube (PMT). QASUME is usually operated in global spectral irradiance mode (Gröbner *et al.* (2005); Hülsen *et al.* (2016)), but during the campaign it was equipped with a collimator tube with a full opening angle of 2.5° allowing the measurement of direct solar spectral irradiance (Gröbner *et al.* (2017)). The measurement range of QASUME during the campaign was limited to 250 nm – 500 nm with a step interval of 0.25 nm, so that one spectrum was measured every 15 minutes. To ensure stable outdoor measurements, the double-monochromator of QASUME was mounted inside a temperature-controlled weather-proofed box (Hülsen *et al.* (2016))."

TECHNICAL CORRECTIONS

"**p. 1 line 2**, "directional irradiance": do you mean "direct irradiance"?"

Authors: Yes, we revise the sentence as: "We demonstrate a Monte Carlo model to estimate the uncertainties of total ozone column (TOC), derived from ground-based direct solar spectral irradiance measurements."

"**p. 1 line 2**, "correlations in the spectral irradiance data" is too generic: do you mean correlation of data within the same spectrum and measured at different wavelengths?"

Authors: We agree that the sentence is too generic. We revise the sentence as: "The model estimates the effect of possible systematic spectral deviations in the solar irradiance spectra on the uncertainties in TOC retrieved."

"**p. 1 line 20**, "analyse uncertainties": uncertainty of what? I guess in ozone retrievals...;"

Authors: Corrected according to reviewer's suggestion. We revise the sentence as: "… data and analyse uncertainties in ozone retrievals for three different spectroradiometers used …"

"**p. 2 line 1-2**: is the order in which the instruments are described (QASUME, AVODOR and BTS) the same as in the abstract (high-end scanning spectroradiometer, high-end array spectroradiometer and roughly adopted instrument)? If not, please avoid confusing the reader;"

Authors: We revise the last paragraph in the introduction as:

"In this paper, we introduce a new method for dealing with possible correlations in spectral irradiance data and analyse uncertainties in ozone retrievals for three different

spectroradiometers used in a recent intercomparison campaign at Izaña, Tenerife, to demonstrate how it can be used in practice. One of the instruments is QASUME (Gröbner *et al.* (2005)) that is the World reference UV spectroradiometer at the World Radiation Center (PMOD/WRC). The second one is an array-based high-quality spectroradiometer BTS2048-UV-S-WP (BTS) from Gigahertz-Optik (Zuber *et al.* (2017a, b)), operated by PTB. The third one is an array-based spectroradiometer AvaSpec-ULS2048LTEC (AVODOR) from Avantes, operated by PMOD/WRC. The field of view of the spectroradiometers has been limited so that they measure direct spectral irradiance of the Sun, excluding most of the indirect radiation from the remainder of the sky."

"**p. 2 line 11**: "total ozone content" is mentioned since the beginning of the paper, but formally defined only in page 6 (Eq. 4). Can you move Eq. 4 a bit earlier?"

Authors: This is true, and thus we move Eq. (4) of our AMT Discussion manuscript to Section 2 so that now it is the first equation of the manuscript.

"**p. 3 line 1**, "vertical profiles were not implemented": what do you mean?"

Authors: We mean that we did not split the atmosphere to horizontal layers at different altitudes and then perform the ozone retrieval fitting. Instead, we use effective atmospheric layers with effective altitudes and temperatures. We revise the sentences in AMT Discussion manuscript as:

"Our ozone retrieval method uses one atmospheric layer to reduce computational complexity. With the one-layer model, the ozone absorption cross-section is a function of the effective temperature, and the relative air mass is a function of the effective altitude of the ozone layer."

"**p. 3 line 2**, "shift the absolute values": values of what? "but should not have an effect": can you justify this hypothesis?"

Authors: PMOD/WRC has a version of the code where the atmosphere is split to horizontal layers at different altitudes and piece-wise calculation of the light transmission is performed. Estimation on possible differences in the TOC uncertainty produced by different layer models will be removed.

"**p. 3 line 10**, "the uncertainties ...are standard deviations": standard deviation of what series/samples? Can you mention which kind of measurements were employed?"

Authors: We revise the sentence as: "The uncertainties stated for h_{eff} = 26 km ± 0.5 km and T_{eff} = 228 K ± 1 K are standard deviations estimated from the vertical profiles in Fig. 1, measured during the campaign on the days 7, 14, 21 Sept. 2016."

"**p. 3 line 11**, "One of the instruments was the QASUME...": already said;"

Authors: We remove this sentence as it was already mentioned.

"**p. 3 line 14**, "every 15 minutes": explain that use of a scanning radiometer involves slower measurements, and why;"

Authors: We include a new paragraph in Section 2:

"The data sets measured by three different spectroradiometers were studied in this work. These spectroradiometers use different techniques to measure the spectral distribution of radiation. Monochromator-based spectroradiometers like QASUME, measure one nearly monochromatic wavelength band at a time, and thus measuring the full spectrum is relatively slow. On the other hand, they usually have significantly better stray light properties than array-based spectroradiometers like BTS and AVODOR, which image the full spectrum at once by dispersing the incoming radiation towards a photodiode array."

We also give measurement intervals of BTS and AVODOR.

"**p. 3 line 15**: what is the spectral range of AVODOR? It is legitimate to say that an instrument "has been corrected"?"

Authors: Regarding the former comment, we revise the text in Section 2 as:

"AVODOR spectroradiometer has a stationary grating and a back-thinned cooled CCD array detector in a Czerny-Turner configuration. AVODOR measures the spectrum from 200 nm to 540 nm with a step size of 0.14 nm in the UV region. During the ATMOZ campaign, the field of view of AVODOR was limited to 1.5° by a commercial collimator tube used, J1004-SMA by CMS Ing.Dr.Schreder GmbH. The spectral range of AVODOR was limited between 295 nm and 345 nm by a combination of two solar blind filters to reduce stray light from the visible and infrared parts of the solar spectrum. The solar blind filters were mounted between the collimator tube and the fiber entrance of the spectroradiometer. One spectrum was measured every 30 seconds."

Regarding the latter comment, we revise the sentence as:

"To measure direct solar spectrum, BTS was equipped with a collimator tube with a full opening angle of 2.8° designed by PTB, and it uses an internal filter wheel system with 8 filter positions together with a specific measurement routine to reduce stray light."

"**p. 4 Table 1**: is this table useful? The same numbers are repeated in Table 4. Also, provide bibliographic references about how each term in the "Standard uncertainty" column was calculated;"

Authors: We move Tables 1, 2, and 3 in the AMT Discussion manuscript to Section 4. We keep Table 1 although some of its data are repeated in Table 4, as its last row states the combined $k = 1$ uncertainty of the spectral measurement. It also makes the describing text clearer when all spectroradiometers are described in similar tables.

We also provide description how standard uncertainties in Tables 1 – 3 are obtained:

"The uncertainties due to radiometric calibration include factors such as the uncertainty of the standard lamp used, and the additional uncertainty due to noise and alignment. QASUME has been validated using various methods, thus the uncertainty due to calibration is low (Hülsen *et al.* (2016)). For QASUME and BTS, we assume the correlations to be equally distributed between *full* correlation, *unfavourable* correlation, and *random* correlation (Kärhä *et al.* (2018)). Spectra measured with AVODOR are significantly noisier, thus half of the uncertainty is associated to the *random* component. Values for instability of the calibration lamp are based on long-term monitoring. The lamp irradiances have been noted to gradually drop with no significant wavelength dependent structure within the wavelength region concerned. Non-

linearity values are estimations of the operators of the devices. Non-linearity is typically manifested so that the responsivity of the device changes gradually from high readings to low readings. This can cause significant change in the TOC values, thus we assume the correlation type to be *unfavourable*. Uncertainties due to device stability and temperature dependence are based on long-term monitoring. The changes have been found to be independent on wavelength in the region concerned, thus *full* correlation is assumed. Noise is the average standard deviation of typical measurements at noon over the wavelength region concerned. The wavelength scales of the devices have been checked using emission lines of gas discharge lamps. The uncertainty values given are the estimated standard deviations of the possible remaining errors after corrections. Wavelength error can introduce a significant change in TOC, because it introduces an error in the form of the derivative of the spectral irradiance. Thus, *unfavourable* correlation is assumed. Most of the uncertainty components are slightly wavelength dependent but to simplify simulations, average uncertainty values are used over the wavelength range between 300 nm and 340 nm."

Regarding the paragraph above, one reference was updated in the AMT Discussion manuscript:

Kärhä P., Vaskuri A., Pulli T., and Ikonen E.: Key comparison CCPR-K1.a as an interlaboratory comparison of correlated color temperature, J. Phys.: Conf. Ser., 972, 012012, 2018. doi:10.1088/1742-6596/972/1/012012

To clarify the uncertainty components in Table 4, we also include new sentences in the AMT Discussion manuscript before line 5 on page 14:

"For components (a) – (d) in Table 4, the mechanism of contributing to the uncertainty of TOC is known. We know the standard uncertainty of the O_3 layer altitude of 26 km to be u = 0.5 km, so we vary the altitude accordingly and note the variance of the resulting TOC."

"**p. 5 line 2**, "fitting the ozone retrieval": can a single quantity be fitted?"

Authors: We revise the sentence as: "They are needed when fitting the spectra at the Earth surface modelled with the ozone retrieval to the measured spectra."

"**p. 5 line 5**, "affects uncertainties with a factor of sqrt(N)": do you mean 1/sqrt(N)? If so, why the Brewer - which measures the irradiance for the ozone retrieval at only 4 wavelengths - is considered a reference in the paper?"

Authors: We revise the sentence as: "In our full spectrum TOC retrieval, the number of data points N which is smaller with a larger wavelength step interval, affects uncertainties with a factor of $1/\sqrt{N}$ (Kärhä *et al.* (2017b); Poikonen *et al.* (2009))."

We include a new reference in the AMT Discussion manuscript:

Poikonen T., Kärhä P., Manninen P., Manoocheri F., and Ikonen E.: Uncertainty analysis of photometer quality factor f_1', Metrologia, 46, 75–80, 2009.

Regarding the latter question, our full spectrum retrieval method does "averaging" in the wavelength domain, whereas Brewer spectrophotometer does it in the time domain. Brewer can measure up to tens of seconds to get millions of photons, so that the photon noise reduces to a level of 0.1%. At this low noise level, it is not critical that only four wavelengths

are used. Averaging over multiple measurement sequences, regardless of the retrieval method used, reduces the noise in TOC by a factor of $1/\sqrt{N_r}$, where N_r is the number of repetitions.

"**p. 6 line 1-2**: give credit to Bouguer, Lambert and Beer (not Huber et al. 1995);"

Authors: We include references to Bouguer, Lambert, and Beer. We also have to acknowledge Huber *et al.*, as the complete ozone retrieval using spectral irradiance measurements and least-squares fitting method have been documented in that paper, and it is also one of the most useful references of our manuscript. Thus, we revised the beginning of Section 3 as:

"In this study, we use an atmospheric ozone retrieval algorithm in many aspects similar to the article by Huber *et al.* (1995). The relationship between the spectral irradiance $E_s(\lambda)$ at the Earth surface and the extra-terrestrial solar spectrum $E_{\text{ext}}(\lambda)$ is based on Beer-Lambert-Bouguer absorption law (Beer (1852); Lambert (1760); Bouguer (1729)) as …"

New references are included in the AMT Discussion manuscript:

Beer A.: Bestimmung der Absorption des rothen Lichts in farbigen Flüssigkeiten, Annalen der Physik und Chemie., 86, 78–88, 1852.

Bouguer P.: Essai d'optique sur la gradation de la lumière (Paris, France: Claude Jombert, 1729) pp. 16–22.

Lambert J. H.: Photometria sive de mensura et gradibus luminis, colorum et umbrae (Augsburg ("Augusta Vindelicorum"), Germany: Eberhardt Klett, 1760).

"**p. 6 Eq. 3**: a reference to the used extraterrestrial spectrum (QASUME-FTS) should be mentioned just after the equation;"

Authors: Corrected according to reviewer's suggestion. We include a new sentence in our AMT Discussion manuscript after line 6 on page 6:

"The QASUME-FTS data set by Gröbner *et al.* (2017) was used as the extra-terrestrial spectrum $E_{\text{ext}}(\lambda)$."

"**p. 6 line 12**: theta is the angle at the observing site, not the angle between vacuum-to air interface;"

Authors: Yes, it is the angle at the observing site. Fortunately, this was a mistake only in the manuscript. It was correctly used in the code. First, effect of the solar zenith angle θ_v at the vacuum-to-air interface at the effective altitude h_{eff} on the relative air mass m is defined as

$$m = \frac{1}{\cos(\theta_v)}. \quad \text{(1R)}$$

After taking the Earth curvature into account, we obtain the relative air mass dependence on the solar zenith angle θ at the observing site as

$$m = \frac{1}{\cos\left[\arcsin\left(\frac{R}{R+h_{\text{eff}}}\cdot\sin\theta\right)\right]}, \quad \text{(2R)}$$

where R is the Earth radius.

"**p. 7 line 10**: the extinction coefficient is defined as dTau/dz, thus it has nothing to do with beta;"

Authors: Corrected according to reviewer's suggestion. Now, we call β as the Ångström turbidity coefficient.

"**p. 7 line 11**: avoid the expression "terrestrial spectrum", the radiation is from the sun, not from the Earth. Use "solar spectrum at the Earth surface";"

Authors: Corrected according to reviewer's suggestion.

"**p. 7 line 22**, "As can be seen, the signal-to-noise ratios ... differ": how can I see it from the figure, without any explanation?"

Authors: It is true that the baseline noise was not clearly shown in Fig. 3 in our AMT Discussion manuscript due to the scaling of the axes. We replace Fig. 3 with Fig. 1R below and revise the paragraph referring to this figure. The new paragraph reads:

*"*Figure 3 presents examples of measurements and modelled values for the spectroradiometers used in this work. As can be seen, the signal-to-noise ratios of the devices differ significantly among different spectroradiometers. All spectra measured by QASUME spectroradiometer are excellent above 10^{-6} $Wm^{-2}nm^{-1}$ with the dynamic range of approximately four orders of magnitude. The dynamic range for BTS is approximately two orders of magnitude and for AVODOR it is less than two orders of magnitude. Based on the analysis in Appendix A, we use absolute least squares minimisation in TOC estimation with $w(\lambda) = 1$ for BTS and AVODOR as it is not affected by the lowest irradiance levels where the stray light and noise are dominant. For QASUME, we use the relative least squares minimisation with $w(\lambda) = E(\lambda)^{-2}$ as it has been used in the past for monochromator-based spectroradiometers, e.g. by Huber *et al.* (1995). The shortest modelling wavelength for the spectroradiometers in this work was set to 300 nm since the typical stray light corrections are not effective below 300 nm (Nevas *et al.* (2014)). The upper wavelength limit was set to 340 nm with all three spectroradiometers as the ozone absorption is not effective above that wavelength."

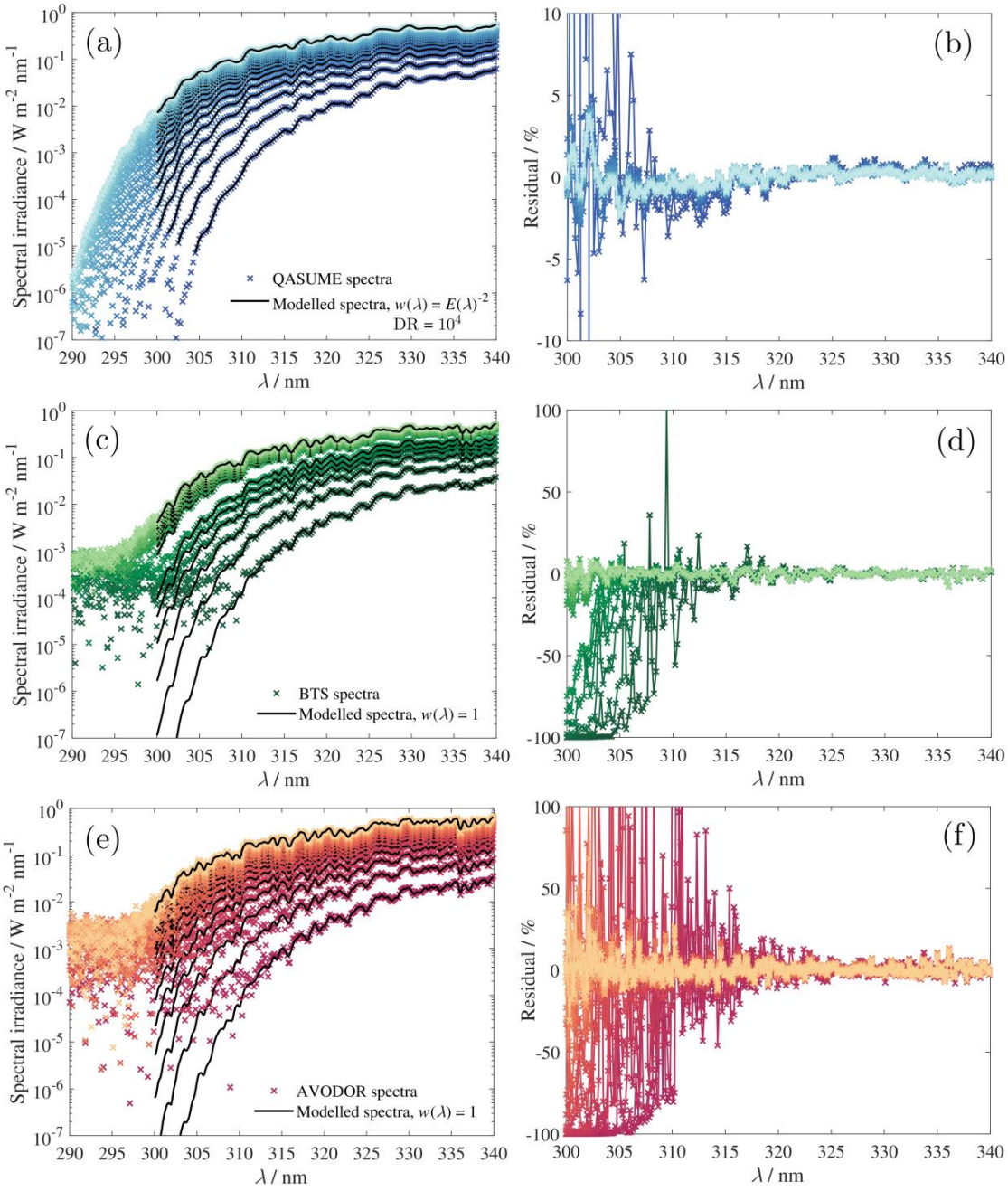


Figure 1R. Examples of fitting the atmospheric model to the direct ground-based solar UV spectra between 300 nm and 340 nm for QASUME (a–b), BTS (c–d) and AVODOR (e–f). In figures on the left hand side, the coloured symbols indicate measured spectra, and the black solid curves indicate modelled spectra. Figures on the right hand side show the relative spectral residuals of the fits. In (a), the abbreviation DR refers to the dynamic range of QASUME data used in the least squares fitting.

"**p. 9 line 3**: "noise" usually defines a random variable, while stray light is a systematic effect. Don't put them together in the same sentence;"

Authors: The new data analysis using absolute least squares fitting for AVODOR and BTS makes this sentence obsolete. The whole paragraph will be rewritten.

"**p. 10, Eq. 11**: define "u";"

> ***Authors:*** We include a following text after line 5 on page 10: "… where $u(\lambda)$ is the relative standard uncertainty of the spectrum."

"**p. 10 line 20**, "does not have any internal limitation to the shape of the error function": what do you mean?"

> ***Authors:*** The MC uncertainty model does not assume any particular spectral shape of the deviation. Fourier series has a property that in its full form it can produce any shape of deviation. This takes place at the Nyquist criterion, where *N* is equal to half the number of wavelengths available. Also with smaller values of *N*, the MC parameters can account for unknown spectral shapes of lower complexity.

"**p. 11 line 3**, "components stating fractions": what do you mean?"

> ***Authors:*** We revise the sentence as: "The uncertainty components divided to the three correlation types have been analysed with the new model. The other components (a)–(d) have been solved using traditional MC modelling because the mechanism for the uncertainty propagating to TOC is known."

"**p. 12 Table 4**: why is "X" used instead of "TOC"?"

> ***Authors:*** We replace "$\tau_X m_X$" with "in exponent" in Table 4 to avoid confusion.

"**p. 12 line 2**: was "r" defined?"

> ***Authors:*** Yes, as there is a sentence before Eq. (12): "Division of the correlation to the three categories introduced are stated for each row as fractions r_{full}, r_{unfav}, and r_{random}."

"**p. 13 line 20**, "a wavelength shift will introduce unfavourable correlations": why? The ozone cross section has a complex shape;"

> ***Authors:*** The ozone absorption cross-section has a complicated shape and the spectral deviation due to the wavelength shift is then also complicated. In other words, the uncertainty due to this complex correlation causes higher uncertainty than noise could cause. Thus, we assume *unfavourable* case of correlation.

"**p. 13 lines 24-25**, "the wavelength shift... should be corrected for the extraterrestrial spectrum": or vice-versa?"

> ***Authors:*** We revised the sentence as "the wavelength shift ... should be corrected **from** the extra-terrestrial spectrum".

March 26, 2018

Response letter to Referee #3 on the manuscript "Monte Carlo method for determining uncertainty of total ozone derived from direct solar irradiance spectra: Application to Izaña results"

Authors: Anna Vaskuri, Petri Kärhä, Luca Egli, Julian Gröbner, and Erkki Ikonen

Article reference: amt-2017-403

Authors: We thank Anonymous Referee #3 for the valuable comments that helped us in improving the manuscript. We have included below our detailed responses to all comments.

Specific comments

"P.4. Table 1. – 3. The table for Uncertainties including the fraction of correlation need further explaining, especially since the issue of correlation is introduced much later in section 4. If these uncertainties have been published in that form earlier, quotation would be helpful in the figure caption. If not, these tables should be moved to section 4 prior to table 4 or in combination with table 4 since the explaining is done on p.15"

> Authors: We agree, and move Tables 1, 2 and 3 to Section 4. Most of the uncertainties are estimated for this manuscript or obtained from personal communication, and they have not been published elsewhere. We include a new paragraph in Section 4 to explain how the uncertainties are obtained:
>
> "The uncertainties due to radiometric calibration include factors such as the uncertainty of the standard lamp used, and the additional uncertainty due to noise and alignment. QASUME has been validated using various methods, thus the uncertainty due to calibration is low (Hülsen et al. (2016)). For QASUME and BTS, we assume the correlations to be equally distributed between full correlation, unfavourable correlation, and random correlation (Kärhä et al. (2018)). Spectra measured with AVODOR are significantly noisier, thus half of the uncertainty is associated to the random component. Values for instability of the calibration lamp are based on long-term monitoring. The lamp irradiances have been noted to gradually drop with no significant wavelength structure within the wavelength region concerned. Non-linearity values are estimations of the operators of the devices. Non-linearity is typically manifested so that the responsivity of the device changes gradually from high readings to low readings. This can cause significant change in the TOC values, thus we assume the correlation type to be unfavourable. Uncertainties due to device stability and temperature dependence are based on long-term monitoring. The changes have been found to be independent on wavelength in the region concerned, thus full correlation is assumed. Noise is the average standard deviation of typical measurements at noon over the wavelength region concerned. The wavelength scales of the devices have been checked using emission lines of gas discharge lamps. The uncertainty values given are the estimated standard deviations of the possible remaining errors after corrections. Wavelength error can introduce a significant change in TOC, because it introduces an error in the form of the derivative of the spectral irradiance. Thus, unfavourable correlation is assumed. Most of the uncertainty components are slightly wavelength dependent but to simplify

simulations, average uncertainty values are used over the wavelength range between 300 nm and 340 nm."

A reference is updated:

Kärhä P., Vaskuri A., Pulli T., and Ikonen E.: Key comparison CCPR-K1.a as an interlaboratory comparison of correlated color temperature, J. Phys.: Conf. Ser., 972, 012012, 2018. doi:10.1088/1742-6596/972/1/012012

"L4-8. Why is uncertainty of radiometric calibration of AVODOR so much higher than for the other instruments? Just because of low SNR in the UV region?"

Authors: Yes, the uncertainty is higher mainly because of the noise. AVODOR response is noisier and less stable as compared with QASUME and BTS.

"The stated uncertainties given in the tables are valid for the whole spectral range of the instrument? I would expect uncertainties related to radiometric calibration and measurement noise to be wavelength dependent. Or are these stated values the upper limit of uncertainties? In the following text, there is a lot discussion about straylight effects. However, there is no explicit uncertainty component related to straylight or straylight correction?"

Authors: It is true that the uncertainties and noise levels are slightly dependent on the wavelength. We simplified the simulations by using average values. A sentence on this is included to the text in Section 4: "Most of the uncertainty components are slightly wavelength dependent but to simplify simulations, average uncertainty values are used over the wavelength range between 300 nm and 340 nm."

Stray light level is difficult to estimate, and to calculate its effect on the TOC uncertainties, we would need the stray light matrix for each instrument (Nevas et al. (2014)) that we do not have. A more practical way of obtaining its effect is to compare the results with a device that does not suffer from stray light. In this case, QASUME is significantly better with respect to stray light than the two array spectroradiometers, BTS and AVODOR. We include discussion about this in the paper.

In the case of array spectroradiometers, the stray light and baseline noise severely affect the TOC analysis at large zenith angles when the fitting is performed with the relative least squares method. This can be seen in Fig. 1R as an inverse U-shape of TOC values determined from BTS and AVODOR spectra when the relative least squares minimisation

$$S = \sum_{i=1}^{n} w(\lambda_i)[E_s(\lambda_i) - E(\lambda_i)]^2 \quad (1R)$$

with $w(\lambda) = E(\lambda)^{-2}$ is used. If the least squares method is modified so that absolute residuals ($w(\lambda) = 1$) are minimised, the inverse U-shape of TOC results at large zenith angles diminishes. With the absolute least squares fitting, the TOC values at noon are 2 DU lower as compared with those estimated using the relative least squares fitting. We include this analysis and Fig. 1R in the revised manuscript. This issue is discussed in more detail in our response to Referee #2.

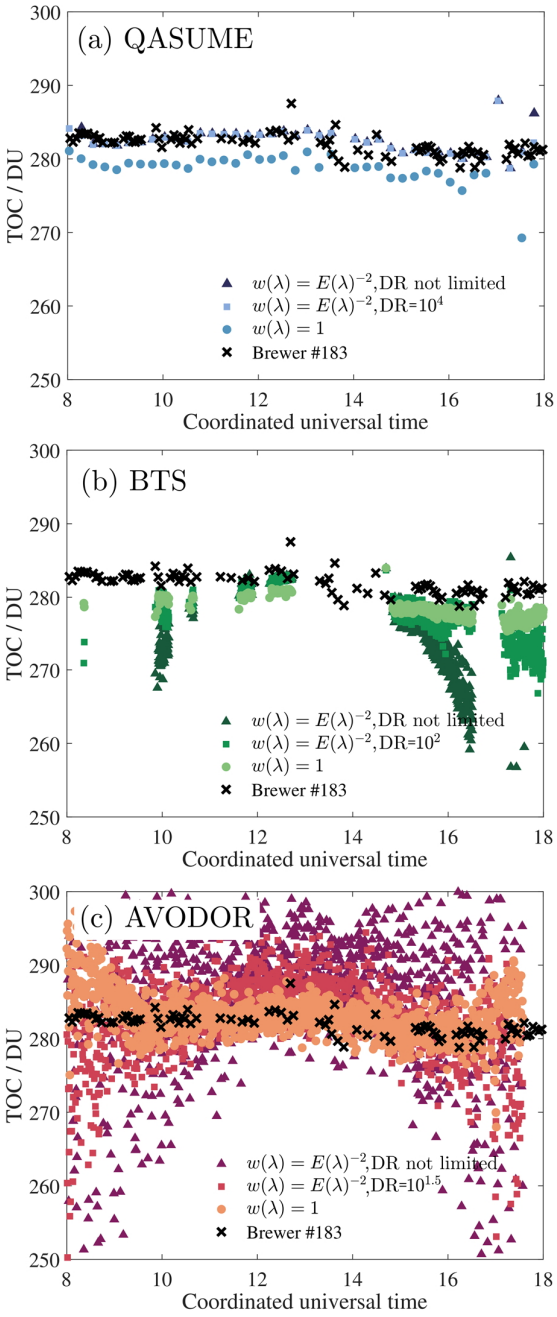


Figure 1R. TOC values estimated using different weighting in least squares minimisation and using Ångström AOD model for QASUME (a), BTS (b) and AVODOR (c). TOC values for Brewer #183 are plotted as black crosses for comparison. The colour codes and the associated figure legends denote the weighting used and the dynamic range DR used.

Reference:

Nevas S., Gröbner J., Egli L., and Blumthaler M.: Stray light correction of array spectroradiometers for solar UV measurements, Appl. Opt., 53, 4313–4319, 2014.

"P. 7, L15 The least square fitting might lead to local minima instead of the global minimum. How is that accounted for? Especially if AOD and TOC are both fitting parameters."

Authors: We performed the least squares fitting using a Matlab function 'lsqnonlin'. We set optimoptions(@lsqnonlin, 'Algorithm', 'trust-region-reflective', 'MaxFunEvals', 1000, 'MaxIter', 1000, 'StepTolerance', 1e-10, 'OptimalityTolerance', 1e-10, 'FunctionTolerance', 1e-10). In our case, one of these tolerances was met by less than 50 iterations when the initial guess values were set to TOC = 200, β = 0, and c = 1.

We included an offset factor c as free fitting parameter in front of $E_{\mathrm{ext}}(\lambda)$ as suggested by Referee #2. After this change, we have three parameters (TOC, β, c) to be fitted. We also set $\beta \geq 0$ using the bound constraints of 'lsqnonlin' function, as aerosols attenuate solar UV spectrum.

To our understanding, global minimum should be achieved quite easily with two or three free parameters, but of course, it depends on algorithms used and the data set to be fitted. We tested the robustness of our retrieval by varying the initial guess values of TOC, β, and c. We varied the initial TOC value between 10 DU – 700 DU, the initial guess value of β between 0 – 0.5, and the initial guess value of c between 0 – 100. Using the initial guess values within such ranges, the free parameters always converged to the same final values and they were independent on the initial guess values. We include text on these tests in the AMT Discussion manuscript on page 7 after line 16.

"P.9, L20 "In this paper, we do this for all components, where the mechanism of contributing to the uncertainty of TOC is known." I guess these components are those with correlation "full" and "random"? This could be specified here."

Authors: Actually, the text refers to those components, where fractions for correlations are not listed in Table 4. Instead, they are labelled as (a) – (d). We clarify this by including a sentence before line 5 on page 14:

"For components (a) – (d) in Table 4, the mechanism of contributing to the uncertainty of TOC is known. We know the standard uncertainty of the O_3 layer altitude of 26 km to be u = 0.5 km, so we vary the altitude accordingly and note the variance of the resulting TOC."

"P.11, Figure 4 That figure is a bit confusing. For underlining the statement on full, unfavorable and random correlation the display of one colored graph is sufficient. The additional information gained from the u=5% and u=2.5% graph, as well as the black solid lines is not explained in the plain text and incomprehensible explained in the figure caption."

Authors: All those curves were intended to show the scalability of the model, but we agree that they were not sufficiently explained in the text. To avoid confusion, we replace Fig. 4 of our AMT Discussion manuscript with Fig. 2R shown below. The new Fig. 4 includes the values obtained by varying the spectral irradiance (circles) and also the values obtained by varying the ozone absorption cross section (crosses). The uncertainty behaves differently if the parameter to be analysed is in front of the equation as a multiplier (such as spectral irradiance) or in the exponent as a direct multiplier to TOC (such as the O_3 cross section). We include a new sentence on page 10, line 16, of the AMT Discussion manuscript as:

"The uncertainty in TOC arising from the spectral deviation in $\alpha_{O_3}(\lambda, T_{\mathrm{eff}})$ is plotted as crosses in Fig. 4 as a function of increasing N."

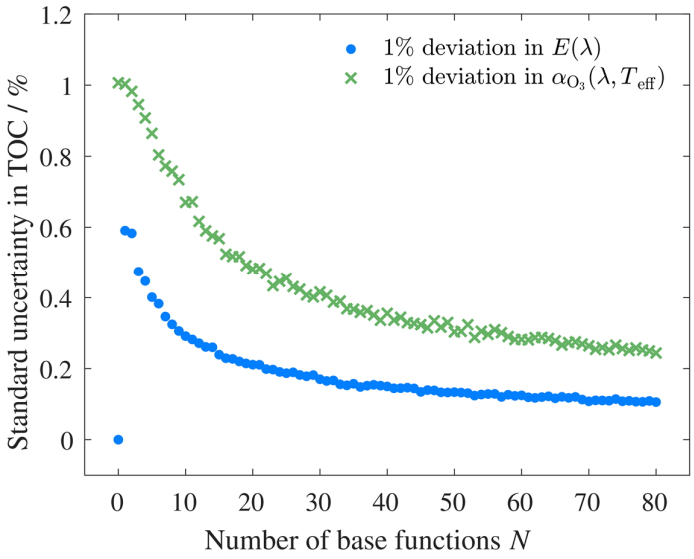


Figure 2R. Standard uncertainties of TOC at local noon as a function of the order of complexity N for QASUME spectroradiometer, with 1% standard deviation in spectral irradiance $E(\lambda)$ plotted as circles, and 1% standard deviation in ozone absorption cross-section $\alpha_{O_3}(\lambda, T_{\text{eff}})$ plotted as crosses.

Additional notes by the authors

As we modified the retrieval algorithm by including a new offset factor c to compensate for full spectral deviations, the results as compared to the AMT Discussion paper will slightly change. Please, see separate response letter addressed to Referee #2 for full details.

March 24, 2018

Response letter to Referee #4 on the manuscript "Monte Carlo method for determining uncertainty of total ozone derived from direct solar irradiance spectra: Application to Izaña results"

Authors: Anna Vaskuri, Petri Kärhä, Luca Egli, Julian Gröbner, and Erkki Ikonen

Article reference: amt-2017-403

Authors: We thank Anonymous Referee #4 for the valuable comments that helped us in improving the manuscript. We have included below our detailed responses to all comments.

Abstract

"**P1, L1** replace "calculate" with "estimate"."

> ***Authors:*** Corrected according to reviewer's suggestion. The new sentence is: "We demonstrate a Monte Carlo model to estimate the uncertainties of total ozone column (TOC), derived from ground-based direct solar spectral irradiance measurements."

Introduction

"P1, L13-14 At this point the authors should make clear that they are talking for correlations in spectral measurements. According to the authors this is the main problem solved when the new methodology is applied. I also suggest adding more information here to help the reader understand what they mean when they refer to "correlations"."

> ***Authors:*** We revise the text about spectral correlations in the introduction and include a new paragraph with some examples about where they might arise:
>
> "TOC can be determined from spectral measurements of direct solar UV irradiance (Huber *et al.* (1995)). We have developed a Monte Carlo (MC) based model to estimate the uncertainties of the derived TOC values. One frequently overlooked problem with uncertainty evaluation is that the spectral data may hide systematic wavelength dependent errors due to unknown correlations (Kärhä *et al.* (2017b, 2018); Gardiner *et al.* (1993)). Omitting possible correlations may lead into underestimated uncertainties for derived quantities, since spectrally varying systematic errors typically produce larger deviations than uncorrelated noise-like variations that traditional uncertainty estimations predict. Complete uncertainty budgets for quantities measured are necessary to understand long term environmental trends, such as changes in the stratospheric ozone concentration (e.g. Molina and Rowland (1974)) and solar UV radiation (e.g. Kerr and McElroy (1993); McKenzie *et al.* (2007)).
>
> Physically, correlations may originate, e.g., from lamps or other light sources used in calibrations. If their temperatures change e.g. due to ageing or current setting, a spectral change in the form of Planck's radiation law is introduced. Non-linearity in the responsivity of a detector causes systematic differences between high and low measured values. The introduced spectrally systematic but unknown changes in irradiance may change the derived TOC values significantly, exceeding the uncertainties calculated assuming that the uncertainty

in irradiance behaves like noise. The presence of correlations in measurements can be seen in many ways. For example, problems have occurred when new ozone absorption cross-sections have been taken into use (Redondas *et al.* (2014); Fragkos *et al.* (2015)). Derived ozone values may change significantly because different systematic errors are included in the different cross-sections. Also, TOC estimated from a measured spectrum often depends on the wavelength region chosen, although the measurement region should not affect the result much."

Regarding these paragraphs, a new reference is included in the manuscript and one reference is updated:

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, 2014.

Kärhä P., Vaskuri A., Pulli T., and Ikonen E.: Key comparison CCPR-K1.a as an interlaboratory comparison of correlated color temperature, J. Phys.: Conf. Ser., 972, 012012, 2018. doi:10.1088/1742-6596/972/1/012012

"P2, L4-5 Do you mean here that the field of view of the spectroradiometers is equal to exactly one solar diameter? If not, then some scattered irradiance also enters the spectrometer."

Authors: The field of view of each spectroradiometer was limited. We revise the sentence as: "The field of view of the spectroradiometers has been limited so that they measure direct spectral irradiance of the Sun, excluding most of the indirect radiation from the remainder of the sky."

We also rename Section 2 as "ATMOZ field measurement campaign and instrument description" and included the field of view of each spectroradiometer in Section 2. The field of view with a full opening angle is 2.5° for QASUME (Gröbner *et al.* (2017)), 2.8° for BTS (Zuber *et al.* (2017b)), and 1.5° for AVODOR according to the manual of the collimator tube used, J1004-SMA by CMS Ing.Dr.Schreder GmbH.

Section 2

"The tables 1, 2 and 3 are presented here without any discussion regarding the presented quantities. I suggest that they should be moved to the uncertainty estimation section (section 4). Furthermore, some discussion (e.g. explaining the presented correlation types, description of how the different uncertainty types were estimated) would be useful."

Authors: We admit these tables are better suited in Section 4, after the spectral correlation types have been introduced. We move measurement uncertainty tables for QASUME, BTS, and AVODOR (old Tables 1, 2 and 3) to Section 4. We also include more discussion about the uncertainty components:

"The uncertainties due to radiometric calibration include factors such as the uncertainty of the standard lamp used, and the additional uncertainty due to noise and alignment. QASUME has been validated using various methods, thus the uncertainty due to calibration is low (Hülsen *et al.* 2016). For QASUME and BTS, we assume the correlations to be equally distributed between full correlation, unfavourable correlation, and random correlation (Kärhä *et al.* 2018). Spectra measured with AVODOR are significantly noisier, thus half of the uncertainty is associated to the random component. Values for instability of the calibration lamp are based

on long-term monitoring. The lamp irradiances have been noted to gradually drop with no significant wavelength structure within the wavelength region concerned. Nonlinearity values are estimations of the operators of the devices. Nonlinearity is typically manifested so that the responsivity of the device changes gradually from high readings to low readings. This can cause significant change in the TOC values, thus we assume the correlation type to be unfavourable. Uncertainties due to device stability and temperature dependence are based on long-term monitoring. The changes have been found to be independent on wavelength in the region concerned, thus full correlation is assumed. Noise is the average standard deviation of typical measurements at noon over the wavelength region concerned. The wavelength scales of the devices have been checked using emission lines of gas discharge lamps. The uncertainty values given are the estimated standard deviations of the possible remaining errors after corrections. Wavelength error can introduce a significant change in TOC, because it introduces an error in the form of the derivative of the spectral irradiance. Thus, unfavourable correlation is assumed. Most of the uncertainty components are slightly wavelength dependent but to simplify simulations, average uncertainty values are used over the wavelength range between 300 nm and 340 nm."

Section 3

"P7, L2 Gröbner and Kerr (2001) did not assume that the air mass factors for aerosols and Rayleigh scattering are equal."

Authors: Indeed, Gröbner and Kerr (2001) did not deal with aerosols. The assumption was taken from the paper by Gröbner *et al.* (2017). We revise the sentence as: "As the ozone and other molecules creating scattering are distributed at different altitudes, we calculate the relative air mass factor m_{R} for Rayleigh scattering at the altitude of 5 km (Gröbner and Kerr (2001)) and approximate the effective altitude of aerosols so that $m_{\mathrm{AOD}} \approx m_{\mathrm{R}}$ (Gröbner *et al.* (2017))."

References:

Gröbner J. and Kerr J. B.: Ground-based determination of the spectral ultraviolet extraterrestrial solar irradiance: Providing a link between space-based and ground-based solar UV measurements, J. Geophys. Res., 106, 7211–7217, 2001.

Gröbner J., Kröger I., Egli L., Hülsen G., Riechelmann S., and Sperfeld P.: The high resolution extra-terrestrial solar spectrum determined from ground-based solar irradiance measurements, Atmos. Meas. Tech., 10, 3375–3383, 2017.

Section 4

"P14, L5 Again, the reference of Gröbner and Kerr (2001) is not correct here."

Authors: We change the reference and revise the sentence as: "Rayleigh scattering and aerosols are set at the altitude of 5 km ± 0.5 km, which influences the relative air mass $m_{\mathrm{R}} \approx m_{\mathrm{AOD}}$ (Gröbner *et al.* (2017))."

Section 5

"P17 Please add more information regarding the linear model used for AOD. E.g., why using the particular model for AOD? Are a and b the same with those of Ångström (1964)? If not, how they are estimated? What happens if the TOC is derived by QASUME and BTS using this linear AOD model?"

Authors: For example, Huber *et al.* (1995) use such a linear model for AOD. The aerosol model by Ångström (1964) can be approximated with a line when a narrow spectral range is modelled. In the linearized AOD model, parameters a and b do not have exact physical meanings, they are just coefficients. Mostly, because of this, we choose to use the model by Ångström (1964), as it is more physical, but we also compare some of our results with results obtained using the linear equation.

In response to the comments by Anonymous Referee #2, we include an offset factor c to the atmospheric model of Eq. (3) of our AMT Discussion manuscript to compensate for full spectral correlations as:

$$E_{\mathrm{s}}(\lambda) = c \cdot E_{\mathrm{ext}}(\lambda) \cdot \exp\left[-\alpha_{\mathrm{O_3}}(\lambda, T_{\mathrm{eff}}) \cdot \mathrm{TOC} \cdot m_{\mathrm{TOC}} - \tau_{\mathrm{R}}(\lambda, P_0, z_0, \phi) \cdot m_{\mathrm{R}} - \tau_{\mathrm{AOD}}(\lambda) \cdot m_{\mathrm{AOD}}\right]. \quad \text{(1R)}$$

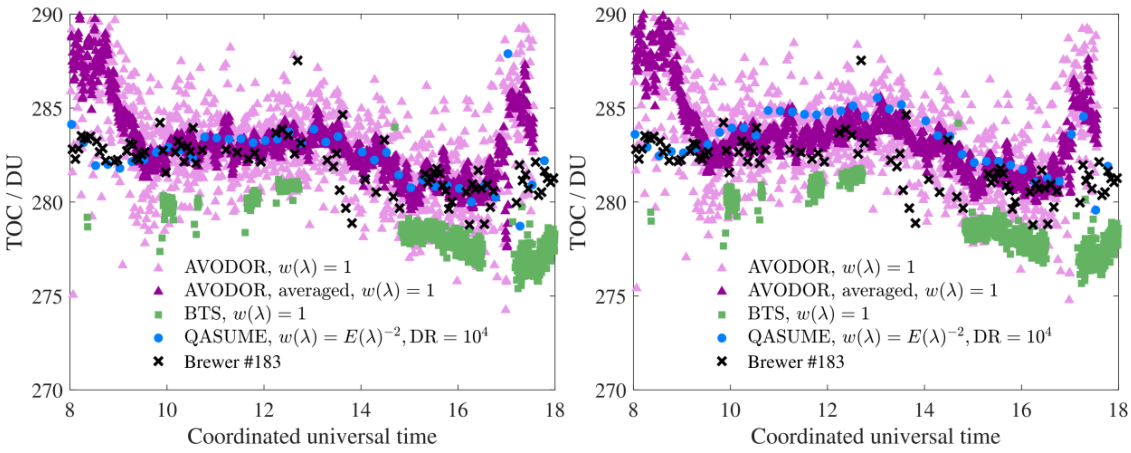
After this change, the atmospheric model has three free fitting parameters: TOC, $\beta \geq 0$, and c, and the TOC estimated for AVODOR at local noon agrees quite well with other instruments. There is still the inverse U-shape in the BTS and AVODOR results, but it diminishes when the relative least squares fitting

$$S = \sum_{i=0}^{N} w_i \left[E_{\mathrm{s}}(\lambda_i) - E(\lambda_i)\right]^2 \quad \text{(2R)}$$

with $w(\lambda) = E(\lambda)^{-2}$ is replaced with the least squares fitting of absolute residuals by setting $w(\lambda) = 1$.

To justify our approach, we compare in Fig. 1R the results obtained using Eq. (1R) with the Ångström AOD model of Eq. (7), to those obtained using Eq. (1R) with the linear AOD model of Eq. (13). When we used linear AOD model, we kept all parameters a, b and TOC as free fitting parameters without bound constraints. In addition, we set $c = 1$, because b produces an almost similar offset factor ($e^{-b \cdot m_{\mathrm{AOD}}}$) as c. Parameter a compensates for slope-like spectral deviations.

According to the new simulations presented in Fig. 1R, the TOC values obtained using the linear AOD model are practically the same for BTS and AVODOR as those obtained using the Ångström model and an offset factor c. With QASUME, the results deviate by 2 DU due to the unconstrained slope factor a of the linear AOD model.



(a) TOC estimated using Eq. (1R) with Ångström AOD model

(b) TOC estimated using Eq. (1R) with linear AOD model

Figure 1R. Absolute TOC values estimated for the spectroradiometers studied. Average of 10 neighbouring values has been included for AVODOR to show the spectral shape behind the noise. Abbreviation DR refers to the dynamic range of QASUME data used in the least squares fitting.

Conclusions

"P18, L27 The results from AVODOR deviate up to 10 DU (and not 10 DU) depending on the SZA. Is the stray light effect enough to explain these discrepancies?"

> ***Authors:*** There is a spectrally constant offset in the spectral irradiance measured by AVODOR that our model in the AMT Discussion manuscript could not handle. We did not take into account how easily full correlations appear in solar UV irradiance measurements, e.g., due to geometrical factors. Thus, we improved our atmospheric model by including an offset factor c as a free parameter in Eq. (1R). Stray light is mostly responsible for the inverse U-shape of TOC (Herman *et al.* (2015)), but using absolute least squares fitting, i.e., by setting the weight to $w(\lambda) = 1$ in Eq. (2R), we get rid of the solar zenith angle dependence. The TOC estimated from the spectra of all the instruments with the improved atmospheric model are presented in Fig. 1R.
>
> We include a new reference in the manuscript:
>
> 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.

"The last paragraph of the conclusions section is now written leads to the conclusion that the main outcome of the study is that the AVODOR is not suitable for TOC measurements, while QASUME and BTS are. In my opinion the main outcome of this study is that the presented method provides more accurate estimations of the uncertainty budget compared to the traditionally used methods. However, it is not adequate for properly estimating uncertainties if the instruments are not characterized for systematic measurement errors. I suggest re-writing the conclusions section in a way that the main conclusions of the study are highlighted."

> ***Authors:*** It is true that the conclusions in its present form give too much weight to the comparison of the devices. We will rewrite the conclusions to give more emphasis to the correlation issues. It is worth noting that in response to Referee #2, we revise the algorithm

for obtaining TOC from spectra (Please, see separate response letter addressed to Referee #2 for full details). The least squares fitting is modified so that the low irradiance values distorted by stray light with BTS and AVODOR get less weight, and the offset of AVODOR gets corrected. After this change, the results are in better agreement, and the daily variation of TOC seen with BTS and AVODOR diminishes (Please, see new results in Fig. 1R). AVODOR seems to work better than first expected. The results are just noisy but quite well in agreement with other devices. We will write the new conclusions after going through all Referee comments. We need to include some discussion about the model change into the conclusions as well, but we try to keep the emphasis on the uncertainty issue.

Additional notes by the authors

As we modified the retrieval algorithm by including a new offset factor c to compensate for full spectral deviations, the results compared to the AMT Discussion paper will change. We will replace Fig. 5 in the AMT Discussion manuscript with Fig. 1R shown in this document.