Response to referees

We thank all three referees for their comments, and their recognition that the present study should open a discussion on improved calculations of the best representative daily value. We have expanded and modified the manuscript in response to all three referees.

Our response to their combined comments is structured below with the original comment in *italic* text, our response in normal text, noting where changes will be made in the revised manuscript with underlined text.

RC1: Anonymous Referee #2

The manuscript presents a new methodology to calculate the best representative daily value (BRDV) of total ozone column measured with Brewer spectrophotometers by using a combination of direct sun and zenith sky observations. The new methodology will clearly have an advantage at some observation sites, because significant ozone variations on a time scale of hours in combination with a limited number of clear sky conditions are better represented with the new method. However, it needs a very good characterisation of the instrument in order to establish the relation between direct sun and zenith sky observations with a small uncertainty.

We agree with the referee that to merge DS and ZS values in this way relies on both measurement types being well-characterised and having small uncertainties. Much effort has been put into improving direct sun measurements and calibrations over the years, with less emphasis on the zenith sky measurement mode. However it should be borne in mind that zenith sky measurements are dependent on the DS mode calibration, and as such, improvements in the determination of these values, and overall instrument stability should lead to improvements in both measurement modes. As a simple safeguard we take the approach of re-determining the ZS polynomials for each inter-calibration period (p3, ll29-30), and in any event, we note that there is the option of applying expanded ZS uncertainties to the relative weightings (p4, ll 29-30).

In general, I think one has to have in mind also, what the purpose of the reported BRDV is. If it is used for validation of satellite estimates of total ozone, then it is more important to focus on the time of the overpass of the satellite. That could reduce the standard deviation of the results, although when using a large number of observations then the average might be unchanged. On p.6, ln. 9-19, this aspect is already discussed, although not shown in detail because 'the results were very similar' to the daily mean values. If the BRDV is used to calculate UV exposure, then the value of the ozone column around local noon would be most significant. Furthermore, the 'daily mean' is anyway biased systematically by the different day length throughout the year, whereas the ozone variations on an hourly time scale are in general independent of daylight or night time.

Our approach was to try and determine a daily value based on the fullest possible set of DS and ZS data available (so to include all valid data, rather than rejecting a potentially large portion). This proposed technique reduces bias away from noon, and so would be a more appropriate measure for daily UV exposure and a more representative value for time series investigations than the current calculation. It is intended as a proposed improvement over the current daily mean value submitted to WOUDC, and users of these values. The referee is correct in that for satellite overpass validation, or high time resolution analysis, individual observations are likely a better choice. Though even here, some choice between DS and ZS measurements would have to be made.

The manuscript is very well written and clearly structured with clear figures. I suggest only one specific point to be clarified by the authors prior to final publication:

p.5, ln. 29: the mean difference between old and new method constitutes 2.79 DU. This should be discussed a bit more detailed as it seems to be significant, due the very high number of samples. It is stated that this is in the order of the 'calibration uncertainty'; does this mean the calibration of the zenith sky relative to the direct sun observations has this significant bias? On the example day shown *in Fig. 1 one can also see such a bias (even considering the standard deviations of the individual measurements). A similar bias is also found in the comparison with satellite data (p. 6, ln. 12/13).*

There are two issues at play here. The ZS polynomials are calculated as a single fit over all cloud conditions present at the site during each inter-calibration period. i.e. considering the mean sky and cloud conditions (overall the fitting procedure gives a bias of only -0.4DU, with a spread between individual DS-ZS pairs of 6.3DU). For a specific day we expect some small DS-ZS biases due to the particular cloud conditions present and those that the referee notes in figure 1 are consistent with the spread of residual DS-ZS values (now noted at p3, ll31-32). The bias of 2.79DU found between traditional and proposed methods, and also exhibited in the comparison to satellite data, instead relates to the new method's increased daily sampling. On this point there are two potential contributions: increased sampling over longer daylight hours where perhaps the TCO is increased, and increased reliance on different internal filters at low solar elevations. In the middle panel of figure 3 there is a suggestion that greater differences are seen during summer when the time span is increased the most (see also figure 2, third row). As to our comment that the value of 2.79DU was of the order of the calibration uncertainty – we referred to the typical spread of values measured by different instruments immediately after calibration (0.5 to 1.0%).

<u>p6, ll24-29: We have expanded and clarified these points in the text.</u>

Overall I think the manuscript is worthwhile to be published in AMT, as it will stimulate an important discussion in the Brewer community. I suggest publication after minor revision.

RC2: Referee #1

This paper is well written and raises a very important question of the daily mean total ozone reporting protocol from the ground-based network of the Brewer spectrophotometers with implications to other type of ground-based instruments. The paper is well written and easy to follow.

The proposed method for calculating the best representative value for daily total column ozone (BRV) using a weighted combination of the direct-sun and the zenith-sky observations will improve the "representative" part of the reported values.

The paper correctly states that the ground-based observations are used in a variety of ways and need to serve the entire ozone monitoring community in the best possible way.

p.1 says that reporting to WOUDC "has traditionally been predicated upon a binomial choice between direct sun and zenith sky observations". This is not quite accurate – there are codes available for many observation types including the focused moon and the ozone values derived from the global irradiance measurements. In fact, WOUDC offers a simple way to register new observation types and report those. Please address this and maybe show how your proposed method can include more than 2 observation types (seems straight forward with weightings).

We thank referee 1 for raising this point. In the current manuscript we concentrate on DS and ZS measurements, as for Manchester and similar mid-latitude sites they are likely to be those relied upon most often. It is true that the methodology can be extended to combining valid observations from any measurement mode, including focussed moon and global irradiance derived total column ozone values. In general we expect that this could be incorporated into WOUDC submissions most simply by registering a new observation type (weighted mean). It would also require guidelines on how it should be derived (for the sake of consistency) with submitters detailing their procedure in the Scientific Support Statement. Alternatively, if stations submit their raw data, or processed individual observations, the weighted mean BRDV calculation could be applied across all sites as a daily summary value.

<u>p1 l9: "binomial" has been altered to "simple". We note that other observation types are possible at p2 ll5-9.</u>

p5 ll4-11: We have added a paragraph noting that the method could be extended to other observation types and its practical implementation in terms of submissions to WOUDC.

The results of applying the proposed method shows some differences compared to the "traditional" method. The paper does provide some comments, but this may require a more detailed discussion. For example, is there a correlation between more ZS only data and high ozone variability during a day that may produce significantly different daily mean depending on when the observations took place?

In an earlier draft we planned to include a figure and more discussion on a closely related point, but we focussed on investigating a correlation between the daily TCO variability and the BRDV differences. There was no clear relation and so it was not included in the submitted version. However we have revisited this aspect, concentrating on the absolute number of contributing DS observations and using the time of year as a coarse proxy for daily variability. There is a clear relation between these parameters: a much greater potential change in the BRDV is exhibited when the number of DS observations nears unity (in line with our assertion in the introduction that partly cloudy days with a small number of DS observations may be most misrepresented). The seasonal effect is fairly weak.

p7 l1-8: We have included an additional panel in figure 3 and discussion at this point in the text.

p1 ll15-16: A sentence has been added to the abstract on this point

The paper makes a good point of suggesting some improvements to the data reporting. What about submission od the individual data points rather than daily averages? Would this not be better for all purposes? WOUDC has a format for this in place.

We agree with referee #1 that reporting individual observations is an option, and would have benefits for certain classes of analysis (satellite validation or detailed studies, for example). For long time series analysis where a consistent approach is needed, or for users who are less well-versed in the details of Brewer operation and data analysis, our view is that a summary value for each day is of benefit to the wider community (see also first response to RC2 on practical implementations).

With minor revision/expansion this paper should be published and will provide a good argument to start a larger discussion on the data submission from the Brewers.

RC3: Anonymous Referee #3

In the manuscript, the authors describe and assess a new methodology for determining a best representative daily value (BRDV) of total column ozone from Brewer spectrophotometer observations. The authors propose a method, which take into account the possibility of changes in the total ozone column (TOC) during the day, and having unevenly sampled observations. All measurements (DS and ZS) of the day are taken into account in case they meet the specified validity criteria and have passed the tail removal check. This is a welcome approach, especially at sites, where cloudiness is frequent and the number of DS measurement is restricted. It would be pity not to use good ZS measurements, as it would be the case if choosing the "traditional" way, where possibly only couple of day's DS measurements would be used. The paper is well written and easy to follow, and is a good opening for discussion in the ozone measurement community. I recommend the paper to be published after minor corrections. Please find here below my specific comments.

Specific comments:

- Please add as reference Karppinen et al. 2016, (especially Figures 2 and 5). Citation: Karppinen, T., Lakkala, K., Karhu, J. M., Heikkinen, P., Kivi, R., and Kyrö, E.: Brewer spectrometer total ozone column measurements in Sodankylä, Geosci. Instrum. Method. Data Syst., 5, 229-239, https://doi.org/10.5194/gi-5-229-2016, 2016.

p5, l6: Karppinen et al (2016) is now referred to in a brief discussion of extending the methodology to other observation types.

- Page 3, line 1: not five wavelengths for ozone measurements?

Five operational wavelengths, plus a dark channel, are used during the instrument observation, though in the standard analysis only four of these operational wavelengths are processed to a final value. We have clarified the text on this point.

<u>p3 l1: The number of wavelengths used has been clarified.</u>

- Chapter 3, Figure 1: Could there be a bias/offset between the DS and ZS measurements of around 5 DU? Looking at quasi simultaneous DS and ZS measurements at around 14h and 16h, there seems to be a difference of around 5 DU, can you explain this?

We expect, and noted, small biases between DS and ZS observations on a particular day. This is attributed to the particular cloud conditions present; the ZS polynomial determination is a single fit over all cloud conditions present at the site during each inter-calibration period. It is for this reason we suggest that one option is to expand the ZS uncertainties at the end of p4. This could be done on a daily basis for days where there are valid DS and ZS measurements.

<u>p4 l30 to p5 l2: We have added the phrase “day-to-day” and an explanatory phrase on this point.</u>

- FIG 1: For days like this, also for cases with evenly spaced DS measurement and changing TOC during the day: why reporting the daily mean? Why not reporting all measurements?

Reporting all the valid measurements is one option. However to provide a consistently determined summary value for each day has benefits for time series analysis or users who wish to know the ozone for secondary analysis. Whether the daily mean value should be provided by users, or by WOUDC by processing of individual measurements is a point for discussion among the community. The former is closer to the current system, but of course the latter would provide a more consistent approach.

- Could also other type of measurements be included in the analyze (FZ, FM)?

Yes, see our first response to RC1.

- Page 3, line 31: What is the range of biases between ds and zs observations?

The overall DS–ZS bias is minimal at –0.4DU. The standard deviation of the population of DS-ZS pairs is 6.3DU.

<u>p3 ll31-32: We have added details of the bias for our dataset here.</u>

- Page 4: increasing the limit value for ZS to 4.0: Does it increase the uncertainty of the measurement?

There will be a small increase in the uncertainty of the BRDV, but this outweighed by the benefit of an increased number of data points and sampling frequency.

- Page 4, line 6: Include a reference to differences in stray light rejection of single and double Brewers.

<u>p4, l7: We have added a reference at the relevant point that notes the differences in stray light rejection between single and double instruments</u>

- Page 4, line 20: How did you end up to choose, for the first (and last) data point, to use the length of the first(last) inter-observation time period? Why not the half of it?

The time weighting for all other data points is from the midpoint of the preceding inter-observation time interval to the midpoint of the following inter-observation time interval. For equally spaced observations, this weighting is equal to the inter-observation time interval. Therefore for the first (and last) data point we chose to weight this equivalently.

What if you have one good ZS measurement in the early morning, no measurements during late morning/midday and many data points (e.g. DS) in the afternoon, isn’t it that in your method the ZS *from the morning will get a big weight as the time t1 in eq. 2. will be big? Can it cause problems, thinking that ZS are not as "good" as DS measurements?*

It is possible to construct sequences of measurements that seemingly cause problems such as this, yes. However, provided care has been taken to determine up to date ZS polynomials, and if necessary, expand the ZS uncertainty, then the authors do not think that the proposed methodology provides a biased result. If in the scenario described there were a strong gradient of ozone, then it would be better to include the morning ZS measurement (provided it is valid and hence has a small uncertainty), than not, when determining the best representative daily value. The referee is correct of course that the long time interval between the ZS and first DS would result in a larger weighting, but this would be offset if it had a larger uncertainty. The first DS will also be given a larger weighting due to the break in valid observations. When there is such a gap in observations, the question is what knowledge do we have about the ozone column during this period? It was this point that lead us to explore the potential benefits of 'Kriging', which would provide a best estimate (and uncertainties) during the gap, based only on the surrounding observations. For many days' observations we obtained good fits, but on occasion spurious results were returned. On balance the weighting proposed was a simpler and more robust route to a BRDV.

- Page 5, What is the min SZA (airmass) in Manchester?

The minimum airmass observed in Manchester is approximately 1.15.

<u>p5 l27-28: We have added the information at this point.</u>

- Page 5: Refer earlier to FIG 2 (actually I didn't find any), otherwise it is difficult to follow the discussion at lines 11-19.

Figure 2 and table 1 now referred to earlier.

<u>p5 l27: Reference to figure 2 added.</u>

- Page 5: I don't find the definition of "representative time", how did you calculate it?

We follow the standard Brewer output, defining representative time as the mean of valid observation times weighted by their TCO values.

<u>P6 ll4-5. This explanation is now added to the text here.</u>

- Page 6: satellite comparison: Is the satellite TOC data meant to represent the daily mean TOC, or is it the overpass time TOC? I don't really see the point to compare the satellite data with other than the nearest of the overpass time. And the satellite should be compared to ground based, not ground-based to satellite. I don't really think that the bias between ground-based and satellite data is due to "that ground-validation of satellites relies upon the traditional methodology", as it can be concluded from your sentence at line 14. But maybe from the way the satellite algorithm is built, satellite instrument errors/uncertainties, etc.

The referee is correct in that it is usually best to compare the satellite TCO to ground-based TCO measured at the overpass time. Here though our aim was to use the satellite record as a common reference to explore how moving from the traditional method to the proposed method of calculating daily means might affect future comparisons between the two instrument sets. In the first part of this comparison we therefore use the mean satellite TCO where there is more than one overpass per day; in the latter part we focussed on selecting DS/ZS measurements close to the satellite overpass time.

<u>p7 l19: added "mean" to clarify satellite data</u>

<u>p6 l22: Line containing quote has been removed</u>

- Page 6: Chapter 5, please add discussion about the problem that e.g. in the morning, the Brewer is looking to East at low solar elevation, and in the afternoon/evening to West at low solar elevation: The

geographical location, to which the TOC is calculated, is not really above the measurement station. In the morning at East from the station, in the Evening at West from the station. What if there is a strong strong East-West ozone gradient? Could the min air mass measurement be after all the most representative for the specific site?

This is an issue that affects any calculation of a daily value, and impacts comparison between ground-based and satellite-retrieved datasets. However, the authors are not persuaded that even in these particular circumstances that the minimum airmass measurement is necessarily an improvement over the proposed method. First if there was a stationary linear gradient as the referee suggests, with e.g. 280DU measured at 9am, 300DU at noon, and 320DU at 3pm, then with equally spaced observations, both the minimum airmass and the suggested method would yield a value of 300DU. If observations were not equally spaced, due to cloud cover, then the traditional method would give a bias, and a minimum airmass measurement may not be available. If the gradient was not linear, then the minimum airmass measurement would be biased according to the overall conditions present through the day. In any event additionally specifying the maximum and minimum would aid in summarising the conditions present during the day, as we mention in Sect. 5.

<u>p5 l13-23: At this point in the text we add a discussion on the problem of sampling along the slant path and how traditional and proposed methods, as well as the minimum airmass measurement, fare.</u>

Technical corrections

- Missing references in the Reference list: Kerr et al., 1981 Smedley et al., 2010

<u>p2 l30: Smedley et al year corrected</u>

<u>p10 ll10-13: Kerr et al reference added.</u>

- Page 3, line 17: "recorded between 1348 and 1549" ->between 13:48 - 15:49 UTC?

<u>p3 l17: Time format corrected.</u>

- Figure 1: Time, UTC ?

<u>p11, fig 1. X-axis label clarified</u>

- Table 1: Explain in the caption what is rep. time (even if we can see it from Figure 2).

<u>p10, table 1: Revised caption now includes full description of “rep. time” and other statistics.</u>

A more representative "best representative value" for daily total column ozone reporting

Andrew R.D. Smedley[1], John S. Rimmer[1], Ann R. Webb[1]

1 Centre for Atmospheric Science, University of Manchester, Manchester, M13 9PL, United Kingdom

Correspondence to: Andrew R.D. Smedley (andrew.smedley@manchester.ac.uk)

Abstract. Long-term trends of total column ozone, assessments of stratospheric ozone recovery and satellite validation are underpinned by a reliance on daily "best representative values" from Brewer spectrophotometers and other ground-based ozone instruments. In turn reporting of these daily total column ozone values to the World Ozone and Ultraviolet Data Centre has traditionally been predicated upon a simple choice between direct sun and zenith sky observations. For mid- and high-latitude monitoring sites impacted by cloud cover we discuss the potential deficiencies of this approach in terms of its rejection of otherwise valid observations and capability to evenly sample throughout the day. A new methodology is proposed that makes full use of all valid direct sun and zenith sky observations, accounting for unevenly spaced observations and their relative uncertainty, to calculate an improved estimate of the daily mean total column ozone. It is demonstrated that this method can increase the number of contributing observations by a factor of 2.5, increases the sampled time span, and reduces the spread of the representative time by half. The largest improvements in the daily mean estimate are seen on days with the smallest number of contributing direct sun observations. No effect on longer-term trends is detected, though for the sample data analysed we observe a mean increase of 2.8 DU (0.82 %) w.r.t. the traditional direct sun vs zenith sky average choice. To complement the new calculation of a best representative value of total column ozone and separate its uncertainty from the spread of observations, we also propose reporting its standard error rather than the standard deviation, together with measures of the full range of values observed.

1 Introduction

Global ground-based monitoring of total column ozone (TCO) relies on the international network of Brewer spectrophotometers since they were first developed in the 1980s (Kerr et al., 1981), which has expanded the number of sites and measurement possibilities from their still-operating predecessor instrument, the Dobson spectrophotometer. Together these networks provide validation of satellite retrieved total column ozone as well as instantaneous point measurements that have value for near real time low ozone alerts, particularly when sited near population centres, as inputs to radiative transfer models at ultraviolet wavelengths, and critically underpin the monitoring requirement of The Vienna Convention for the Protection of the Ozone Layer, 1985.

There are recent indications that for the first time since the treaty was enacted, and CFCs and other depleting substances were banned, the ozone layer is showing signs of recovery (WMO, 2014, and references therein). These and related trend analyses, however, use daily mean TCO values as their starting point. In the parlance of the World Ozone and Ultra-Violet Data Centre (WOUDC), the WMO data centre for ground-based ozone data, the daily values submitted by data originators should be the "best representative values" (WOUDC, 2016). For Brewer spectrophotometers a cascade of choices is recommended as follows. If available, the mean of valid direct sun (DS) measurements is preferred. If no valid DS observations are available for a given day, then the mean of all valid zenith sky (ZS) measurements is used. If no valid DS or ZS observations are available, the last choice is to rely on the mean of valid focused moon observations, a measurement mode predominantly used at high latitude stations. Here we only consider the choice between DS and ZS observations.

The majority of effort spent on calibrating Brewer spectrophotometers is directed towards ensuring high quality DS calibrations are distributed globally through the Brewer reference triad (Fioletov et al., 2005), intercomparisons at the Regional European Calibration Centre and through initiatives such as COST Action ES1207 and EUBrewNet. ZS observations are then linked to the DS calibration through a polynomial fit of quasi-synchronous DS and ZS observations (Kipp & Zonen, 2005). This additional calibration step explains the default preference for DS over ZS measurements as it incurs a small associated uncertainty. However at mid- and high-latitude stations in particular the annual mean cloud fraction can exceed 50 % (Wilson and Jetz, 2016) and limits opportunities for recording viable DS observations. As a consequence, for a high fraction of days the best representative daily value (BRDV) is based upon zenith sky measurements. More crucially during partly cloudy days, the BRDV can be reliant on a small number of individual DS observations (<5), which may be biased towards either the start or end of the day, whilst a greater number of valid ZS observations from throughout the observation period are rejected

This gives rise to the question: could a more representative daily value be obtained from an increased number of ZS measurements than from a small number of DS measurements? To answer this still forces a choice between valid DS and ZS observations as the number of DS measurements falls – whichever set of observations is chosen, a set of otherwise valid data is not incorporated into the calculation of the best representative value. Therefore we propose an alternative methodology to calculate a best daily representative value that retains both direct sun and zenith sky measurements, taking into account their relative uncertainties and periods of time when valid measurements are more frequent.

2 Instrument and data processing description

Brewer spectrophotometers, their operation and standard data processing routes have been described previously in the literature (Brewer, 1973; Fioletov et al., 2005; Smedley et al., 2012; Savastiouk and McElroy, 2005). For context we outline the key points here. The core of each instrument is a single or double monochromator unit whose output is detected by a photomultiplier tube. For the DS measurement mode the input is from a rotating prism assembly pointed towards the sun's disc. Column ozone observations are achieved by rapidly repeated measurements at 5 operational wavelengths over a period of approximately 3 minutes, and a final value calculated by implementing the Lambert-Beer law and knowledge of the absorption cross-section of ozone molecules. ZS observations are made in the same way but the rotating prism is instead directed to collect scattered light from the zenith, and then an empirical polynomial adjustment is applied. This polynomial adjustment assumes that the apparent ozone column from the zenith sky measurement is quadratic in both air mass factor and the actual column ozone. The nine constants necessary are determined from a large number of quasi-simultaneous DS and ZS measurements (>500) and are instrument and site specific. This relationship is usually determined at the instrument's home site, rather than during an intercomparison or calibration exercise, though Fioletov et al (2011) described an improved RT modelling based methodology that reduced the instrument-specific unknowns to two parameters (though nine constants are still necessary). However the polynomial constants are determined, the ZS observation is then found by solving the relevant quadratic equation.

3 A more representative daily value

For a site where direct sun can be guaranteed for the majority of the day, an instrument could be scheduled to only attempt DS observations at regular intervals (together with the necessary diagnostic routines) and the mean of these observations will be a reliable estimate of the actual daily mean TCO overhead at the station. For other sites where cloud is more variable and unpredictable, the observational schedule must contain a combination of both ZS and DS measurements. However local cloud cover conditions may only permit a small number of DS measurements to be successfully recorded. Figure 1 shows an example day where only 4 DS observations were recorded between 13:48 and 15:49 UTC, and where their arithmetic mean (298.23 DU) differs substantially from the daily mean TCO at the station as indicated by ZS observations. In order to avoid the potential binomial choice between a small number of DS observations and a greater number of ZS observations, we propose a weighted daily mean that utilises all valid DS and ZS values.

Our aim is to construct a daily mean that has the following properties. In the absence of either any valid DS or ZS observations it produces the same result as the standard method (once clustering of observations is accounted for). With the addition or subtraction of a single DS data point, there is a graceful change in the BRDV and the overall time period it represents. It should represent as fully as possible the day's TCO observations. It should give equal weighting to equal periods of time and hence account for time clustering of valid observations and for their relative uncertainties. It should be able to be applied to historic data and not necessitate any changes to the instrument's future schedule or data collection routines.

The proposed methodology is as follows. The first pre-requisite is for ZS polynomials to be assessed regularly (here these have been recalculated using the full available dataset for each inter-calibration period) to ensure the individual DS and ZS observations are comparable and there is minimal bias in the ZS observations (for example in our dataset overall DS–ZS bias = –0.4 DU; standard deviation of distribution of individual DS–ZS pairs = 6.3 DU).

All DS and ZS measurements are then filtered to remove those that do not meet the validity criteria. Observations that have a standard deviation of > 2.5 DU for DS and > 4.0 DU for ZS are rejected. We note that the standard choice of standard deviation threshold is 2.5 DU for ZS observations, but increasing the limit to 4 DU does not introduce any bias and increases the total number of valid observations (Fioletov et al., 2005; Fioletov et al., 2011). Observations at air mass factors > 4 are also rejected for single monochromator instruments, but this limit is raised to 6 for double monochromator instruments due to their improved stray light rejection (Karppinen et al., 2015). To ensure that any residual bias is not present at high air mass factors, an additional tail removal step is applied. For this the day's data is smoothed with a 30 minute running average filter, and end periods of time where the smoothed TCO exhibits apparent rates of change > 20 DU/hr are identified. Any observations falling within these periods are then removed.

At this stage the remaining DS and ZS values meet the specified validity criteria, and have passed the additional tail removal check. To form a BRDV from these individual observations, we calculate a weighted mean of the full set of data points, but where the weighting has two components: the time for which the observation is representative, and the uncertainty of each observation, as in Eq. (1):

$$\overline{X} = \frac{\sum_i X_i w_i}{\sum_i w_i} \quad (1)$$

where $\overline{X}$ is the BRDV, X_i are the individual observations (both DS and ZS), and w_i is the weighting for each observation. The weighting is defined in Eq. (2) as:

$$w_i = \frac{t_i}{\sigma_i^2}, \quad (2)$$

where t_i is the time from the midpoint of the preceding inter-observation time interval to the midpoint of the following inter-observation time interval. For the first data point we instead use the length of the first inter-observation time period, and likewise for the last data point. In all cases σ_i are the uncertainties for each individual observation, taken as the normal measurement standard deviation and used as part of the validity test. If no account were taken of the relative uncertainties of each observation, nor of their time intervals, this formulation reduces to the simple arithmetic mean of the valid observations.

We also note that this methodology could be applied to all data acquired without applying a threshold standard deviation validity filter as data points with large errors will contribute to the BRDV proportionally less. However more care needs to be taken as regards relaxing the airmass threshold requirement as small biases may be introduced, inflated by the effect of observing at high airmass, whilst the uncertainty will not have been captured by the intrinsic standard deviation of the observation. Further the ZS uncertainties could be expanded appropriately to account for any day-to-day bias between ZS and

DS observations under differing sky conditions, or alternatively to incorporate the DS-ZS polynomial fit mean residual, for example.

For higher latitude sites where other measurement modes are relied upon, such as focussed moon, focussed sun, or TCO derived from global spectral irradiance, these observations could also be incorporated into the BRDV calculation in a similar way (see seasonal variation of observation types in Karppinen et al, 2016). The prerequisites would be that each individual observation had an associated uncertainty and that the observations from different measurement modes had been homogenised beforehand. In terms of practical implementation, if the method were adopted by the community a new observation type would have to be registered at WOUDC (a mechanism that is already available), with relevant details added to the Scientific Support Statement as necessary. For stations that submit raw data, or processed individual observations, the weighted mean BRDV calculation could be applied across all sites as a daily summary value.

It is also worthwhile considering the strengths of this methodology under specific theoretical conditions. If, for example, on a given day there is a strong linear east-west ozone gradient present, then the most appropriate daily measure should return a value similar to the TCO above the site. The traditional method risks producing a daily value that could be substantially different if, due to cloud cover, only a few valid DS measurements can be recorded during early morning or late in the day when the TCO is being sampled to the west or east of the site. In partly cloudy or cloudy conditions a minimum airmass TCO measurement may not be obtained. In contrast the proposed method guards against these issues as ZS observations could be sampled more fully through the day, whilst the contribution from DS measurements would represent the effective TCO along the slant path when the direct solar beam is visible. As a result the proposed BRDV TCO calculation is more appropriate for UV exposure studies than the traditional calculation, more representative of the conditions throughout the day, and more resilient than relying on a single value at minimum airmass, for example. For non-linear spatial gradients in TCO then limited DS measurements could result in a value more different still from the mean TCO overhead, while the bias from selecting the TCO near minimum airmass would depend on the spatial distribution of ozone.

4 Sample results

To demonstrate the impact of this method on real world data, we apply it to the 2000–2016 data record from Brewer spectrophotometer #172, located in Manchester, UK [53.47°N, 2.23°W] (see table 1 and fig 2). For context the minimum airmass observed at this location during the summer solstice is approximately 1.15, whilst during the winter solstice it is 4.15.

Overall we see an increase in the mean number of contributing observations (N) from 10.72 to 24.06, with the upper 10th percentiles also increasing from 24 to 46. The effect is dominated by summer-time measurements that show an increase of 150 % from 14.68 to 36.68 averaged over the three months bracketing the summer solstice. Whilst the effect is still present during winter months (when data collection is inherently more difficult due to the lower solar elevations), the improvement is smaller: the mean N increasing from 5.71 to 8.54 contributing observations per day.

As expected we see concomitant tightening of the distribution of representative times (defined as the mean of valid observational times, weighted by their TCO values) around solar noon (close to midday). There is also a skewing of the observational time span distribution to longer periods. The representative time is symmetrical about solar noon in both the traditional and new methods, but the width of the annual distribution (defined as the interdecile range) is halved from 4.05 h to 2.01 h, showing the new method results in BRDVs more representative of conditions at solar noon Again due to the longer day length the improvement is accentuated during summer (6.5 h reduced to 2.7 h), but still present during winter months (1.56 h reduced to 1.20 h). Time spans for the whole year are generally skewed to the right-hand side of the distribution, though the upper and lower bounds do not change (being limited by number of daylight hours and instances of single contributing observations respectively). Much of the skewness in the annual distribution is attributable to that occurring during the summer subset (fig 2, third row, second column), where the lower 10th percentile increases from 0.5 h to 10.0 h.

Taken together these results demonstrate that the method enables a more representative daily mean to be calculated, predominantly by sampling more fully through each day and over a wider range of weather conditions. However it is prudent to investigate the impact on the overall time series and trends.

Focussing on the 2006–2016 subset for clarity, in figure 3 we see only a small impact on the monthly mean TCO values from applying the methodology described, with no discernible trend or annual cycle in the difference (fig. 3, upper and middle panels). Regression analysis shows the trend in the difference between traditional and proposed methodologies to be –0.0153 DU/yr, but this is not significant (Kendall-Mann test, p = 0.275). Specifically over the 2006 to 2016 period we find a skewed distribution of daily differences between the new and traditional BRDVs with a mean increase of 2.79 DU, an amount comparable to the calibration uncertainty (the spread of values exhibited by different instruments immediately after calibration), and with the upper and lower 10th percentiles being –2.67 DU and 10.92 DU respectively (equivalent to –0.79 % and +3.22 % of the annual mean TCO). The methodology does not substantially alter the form of the time series with an R^2 coefficient of 0.984 (fig 3, lower middle panel). It should be emphasized that this bias of 2.79DU does not result from the ZS polynomial procedure, nor is it related to any bias between individual DS and ZS measurements: the overall bias of the polynomial fit is only –0.4DU. Instead the bias noted here relates to the new method's increased sampling over longer daylight hours. The underlying cause is likely related to a longer sampling period where the TCO is larger, or to increased reliance on different internal filters. We anticipate that the former is dominant as figure 3 (middle panel) suggests increased differences during summer months when the time span is increased the most.

In figure 3 (lower right panel) we explore the ranges of differences between methodologies further. It is anticipated that the greatest differences between traditional and proposed methods will be seen on days with high ozone variability and a low number of contributing DS measurements (or a high fraction of ZS observations). Plotting the TCO difference against the number of valid DS measurements, the greatest variability is seen for a single valid DS measurement with the distribution rapidly narrowing as the number of valid DS observations increases. We distinguish days according to their seasons, with those from summer and autumn [JJASON] marked in dark grey and winter and spring [DJFMAM] marked in red. Typically TCO exhibits a much larger variability during winter and spring at this location, though the effect is not overly strong and the number of DS observations is a better indicator for a large potential improvement in the BRDV.

Together these results suggest that there should be no impact on long-term trends at a site where the data record is derived from a single instrument type. However, there could be implications where there has been a change in data sampling method. Moving from a semi-manual Dobson spectrophotometer that makes a limited set of observations on a predefined schedule to a Brewer spectrophotometer that operates quasi-continuously and selects a daily value on the traditional DS vs ZS choice, could introduce a small step change due to this affect, which may contribute to a perceived trend. Likewise applying the proposed method to only part of a data record, because individual historical measurements have been lost, for example, could also introduce a small step in the overall record.

Testing the influence of the new methodology in terms of the agreement between ground-based and satellite retrievals (fig. 4), we find a marginal improvement in the ground vs satellite TCO for daily mean data in terms of their R^2 correlation (0.9560 vs 0.9455), and best fit slope (1.0041 vs 0.9930). More noticeably, there is a narrowing of the distribution of differences (ground-minus-satellite) from an interdecile range of 26.86 DU to 24.01 DU, whilst the mean is shifted to higher biases (from 10.49 DU under the traditional DS-ZS preference to 13.28 DU). Satellite retrievals were also compared against the closest individual observation to the overpass time under two assumptions. First, selecting the nearest valid DS measurement as a preference, and if none were available, then selecting the closest ZS observation (equivalent to the traditional DS-ZS choice). Second, selecting the closest observation to the mean overpass time with no preference for observation type (equivalent to the proposed BRDV calculation). The results nonetheless were very similar to those for BRDV in figure 4.

5 Measures of daily spread and estimating real time TCO

While the focus of this study is on the determination of a more representative daily TCO value, there are a number of related issues concerning reporting of the daily spread that will be discussed in this section.

At present WOUDC recommendations are to report, in addition to the best representative value, a standard deviation for the day's observations, which implicitly assumes a normal distribution. Whether a day's observations of the underlying ozone column necessarily falls within a normal distribution is not obvious, nor are the authors aware of any evidence in the literature. To that end we have applied the Kolmogorov-Smirnov test (Massey, 1951) to each day's observations for the same 2006–2016 subset of data. The null hypothesis for this test is that the individual daily observations come from a standard normal distribution and on application we find that this null hypothesis is not rejected for any of the days in our test sample. That is, all can be considered being taken from a normal distribution.

Whilst this result does not undermine the use of the standard deviation as a measure of the spread of the day's data, we propose that other metrics may be more useful. Specifically to separate out the uncertainty in the best representative daily value from the range exhibited by individual observations, a more useful measure would be to use the standard error of the weighted mean to indicate the uncertainty of the best representative value, plus additional metrics relating to the maximum and minimum TCO observed. These latter could be the strict maximum and minimum, or to guard against the influence of short-duration spikes, the upper and lower 10th or 25th percentiles could be used, for example.

Whilst developing the methodology described in Sect. 3, the geostatistics analysis route known as 'Kriging', or Gaussian process regression, was also tested (Bailey and Gatrel, 1995; Lophaven et al., 2002). This analysis produces the best linear unbiased estimator of the actual underlying TCO at times intermediate to the observations, and also produces an associated uncertainty. In brief, it performed well for days where there are a larger number of contributing observations, but showed poorer performance during winter or other days with few observations. This latter issue is in part due to the complex nature of applying the method, where for few observations there is a risk of overfitting. However for studies where short-term prediction of the TCO and its near term uncertainty is of interest, such as real-time estimates or nowcasting, Kriging may find applications. More generally its applications could include spatial analysis and interpolation of TCO and surface irradiance, two fields where global datasets are reliant on a limited number of measurement sites.

6 Conclusions

In this study we propose, describe and assess a new methodology for determining a more representative best daily value of total column ozone from Brewer spectrophotometer observations. This method overcomes the limitations of making the traditional choice between a possibly small number of direct sun measurements and zenith sky measurements. It requires a homogenised set of DS and ZS data as a pre-requisite, but then, by taking a weighted mean and accounting for both the uncertainty associated with each individual observation and the time period the observation represents, produces a more representative value based on the full set of daily observations.

Applying the new method to the 2000–2016 dataset from Brewer 172 stationed at Manchester [53.47° N, 2.23° W], we show that the number of contributing observations is more than doubled from an average of 10.72 to 24.06 per day; increased numbers of observations are found in both summer and winter, though the fractional increase is greater during the summer. Similarly the interdecile range of mean representative times is approximately halved throughout the year, whilst the time span of contributing observations is skewed towards longer hours, predominantly during the summer months. Together these findings demonstrates that the method results in a substantial improvement in sampling and utilisation of valid observations, and hence improves the representativeness of the daily mean TCO. The issue of rejecting otherwise valid data is also removed. We find no evidence of impact on supra-annual trends from applying the new method, though the ground-satellite bias is increased for this station by 2.8 DU. We also note that a change in daily sampling when one instrument type replaces another at a site could contribute to introduce a small step in the data record, and similarly, care should be taken if reprocessing only a partial data record.

To complement our proposed BRDV calculation, we also recommend reporting the standard error of the daily mean value, and replacing the standard deviation by a more complete measure of the daily spread such as the upper and lower limits of the interdecile range, or simply the maximum and minimum observed values.

Data availability

The underlying data used in this study can be accessed at the World Ozone and Ultra-Violet Data Centre (Smedley et al., 2017).

Author contribution

A.R.D. Smedley was primarily responsible for the data collection, processing and monitoring of Brewer spectrophotometer #172 and led the manuscript preparation. J.S. Rimmer assisted with data collection, contributed to the manuscript and securing of funding. A.R. Webb contributed to the manuscript, securing of funding and was Principal Investigator on the overall grants.

Acknowledgements

Stratospheric ozone and spectral UV baseline monitoring in the United Kingdom is supported by DEFRA, The Department for the Environment, Food, and Rural Affairs, since 2003.

References

Bailey, T. C. and Gatrel, A. C.: Interactive spatial data analysis, Addison-Wesley, 1995.

Brewer, A. W.: A replacement for the Dobson spectrophotometer? Pure Appl. Geophys., 106–108, 919–927, doi:10.1007/BF00881042, 1973.

Fioletov, V. E., Kerr, J. B., McElroy, C. T., Wardle, D. I., Savastiouk, V., and Grajnar, T.S.: The Brewer reference triad, Geophys. Res. Lett., 32, L20805, doi:10.1029/2005GL024244, 2005.

Fioletov, V. E., McLinden, C. A., McElroy, C. T., and Savastiouk, V.: New method for deriving total ozone from Brewer zenith sky observations, J. Geophys. Res. Atmos., 116(8), 1–10, doi:10.1029/2010JD015399, 2011.

Karppinen, T., Redondas, A., García, R. D., Lakkala, K., McElroy, C. T. and Kyrö, E.: Compensating for the effects of stray light in single-monochromator Brewer spectrophotometer ozone retrieval, Atmos. - Ocean, 53(1), 66–73, doi: 10.1080/07055900.2013.871499, 2015.

Karppinen, T., Lakkala, K., Karhu, J. M., Heikkinen, P., Kivi, R. and Kyrö, E.: Brewer spectrometer total ozone column measurements in Sodankyl, Geosci. Instrumentation, Methods Data Syst., 5(1), 229–239, doi:10.5194/gi-5-229-2016, 2016.

Kerr, J. B., McElroy, C. T. and Olafson, R. A.: Measurements of ozone with the Brewer spectrophotometer, in Proceedings of the Quadrennial International Ozone Symposium, edited by J. London, pp. 74–79, Natl. Cent. for Atmos. Res., Boulder, Co., 1981.

Kipp & Zonen.: Brewer Mk III spectrophotometer operator’s manual, Delft, 131p, 2005.

Lophaven, S. N., Nielsen, H. B., and Søndergaard, J.: DACE: A MatLab Kriging toolbox, Report IMM-TR-2002-12, Informatics and Mathematical Modelling, DTU, 28 p, 2002. Available at: http://www2.imm.dtu.dk/projects/dace/dace.pdf

Massey, F. J. Jr.: The Kolmogorov-Smirnov test for goodness of fit. J. Am. Stat. Assoc., 46(253), 68–78, 1951.

Savastiouk, V. and McElroy, C.T.: Brewer spectrophotometer total ozone measurements made during the 1998 Middle Atmosphere Nitrogen Trend Assessment (MANTRA) campaign, Atmos.–Ocean, 43, 315–324, doi:10.3137/ao.430403, 2005.

Smedley, A. R. D., Rimmer, J. S., Moore, D., Toumi, R., and Webb A. R.: Total ozone and surface UV trends in the United Kingdom: 1979–2008, Int. J. Climatol., 32, 338–346, doi:10.1002/ joc.2275, 2012.

Smedley, A. R. D., Rimmer, J. S., and Webb, A. R.: UK ozone and UV monitoring site no. 352, WOUDC, http://doi.org/10.17616/R32C87, 2017.

Wilson, A. M. and Jetz, W.: Remotely sensed high-resolution global cloud dynamics for predicting ecosystem and biodiversity distributions, PLoS Biol., 14(3), e1002415, doi:10.1371/journal. pbio.1002415, 2016.

WMO (World Meteorological Organization).: Scientific Assessment of Ozone Depletion: 2014, World Meteorological Organization, Geneva, Switzerland, 416p, 2014.

WOUDC: World Ozone and Ultraviolet Data Centre (WOUDC) Contributor Guide v1.2.2, Environment and Climate Change Canada, 69p, 2016. Available at: http://woudc.org/archive/Documentation/GuideBooks/woudc-contributor-guide-en.pdf

Table 1. Summary statistics for data shown in figure 2: number of contributing observations, representative observation time, and time span of contributing observations. For each case values shown are arithmetic means, and in brackets, lower 10th percentile and upper 10th percentile.

	No. of Observations	Rep. time [h]	Time span [h]
Annual (trad.)	10.72 [2 24]	12.12 [10.19 14.24]	5.15 [0.06 10.94]
Annual (new)	24.06 [6 46]	12.02 [11.01 13.02]	7.61 [2.50 12.35]
Summer [MJJ] (trad.)	14.68 [2 31]	12.22 [9.26 15.56]	7.76 [0.50 12.25]
Summer [MJJ] (new)	36.68 [21 55]	11.99 [10.63 13.33]	11.56 [10.00 13.00]
Winter [NDJ] (trad.)	5.71 [1 11]	12.00 [11.27 12.83]	2.18 [0.00 4.00]
Winter [NDJ] (new)	8.54 [4 16]	12.01 [11.40 12.60]	2.92 [1.50 4.50]

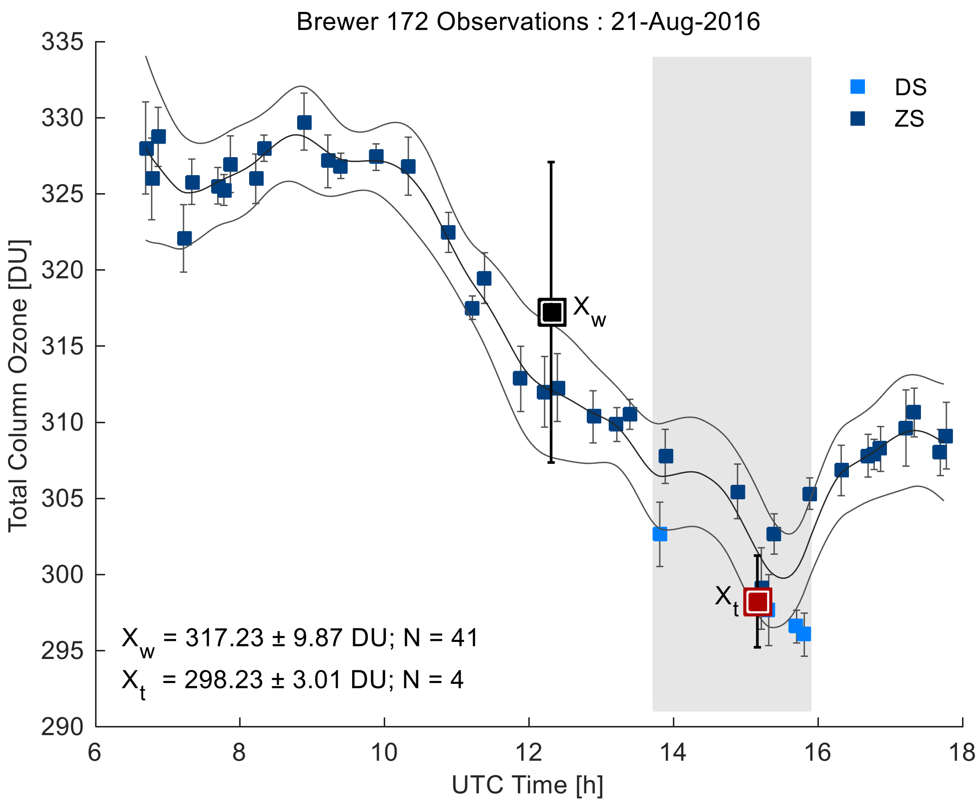

Figure 1. Example day showing valid DS and ZS observations and their standard deviations. Also shown are the daily representative value based on the traditional (arithmetic mean, DS>ZS preference) methodology (red outlined square, X_t), and the daily representative value formed through the method described herein (black outlined square, X_w). The shaded area shows the time coverage for data points contributing to the traditional estimate of the BRDV, and N is the number of contributing observations to each BRDV estimate.

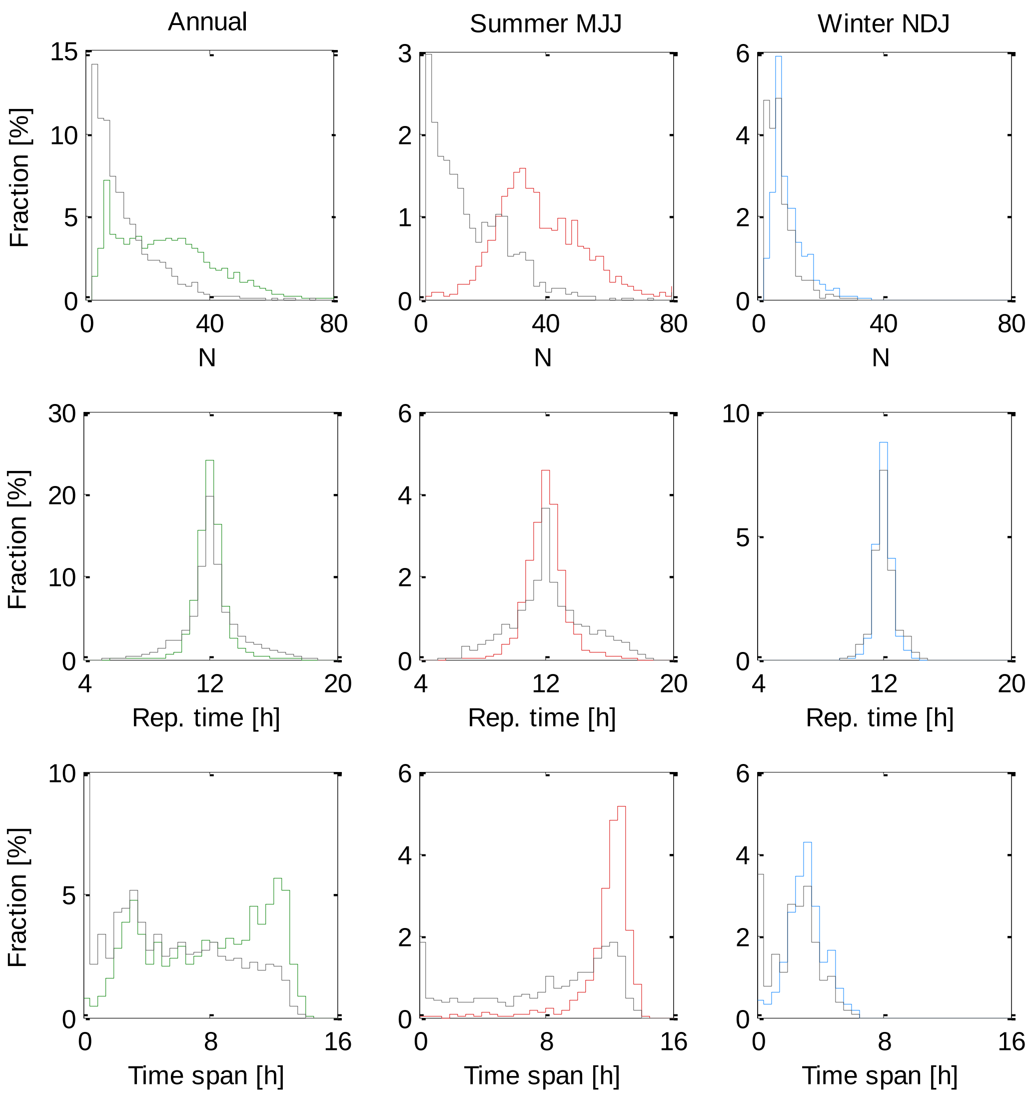

Figure 2. Histograms of number of contributing observations (N, first row), representative observation time (second row), and time span of contributing observations (third row) for 2000–2016 all months (first column), summer months (May-Jun-Jul, second column), and winter months (Nov-Dec-Jan, third column). Grey traces show results from traditional method, coloured traces show results for methodology described in the present study.

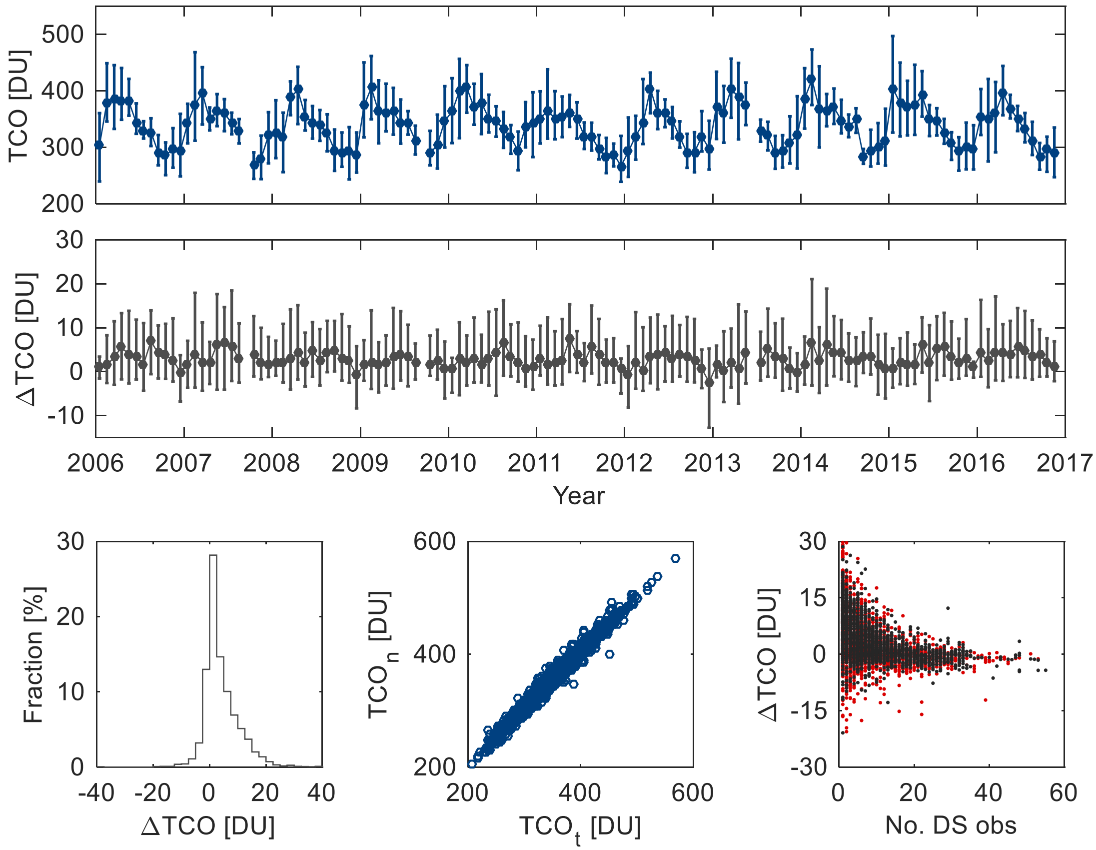

Figure 3. Upper full width panel: Traditional monthly mean TCO; line length shows upper and lower monthly 10th percentiles. Lower full width panel: Mean monthly daily difference between new and traditional best representative values, plus upper and lower monthly 10th percentiles. Lower left panel: Histogram of daily differences between traditional and new daily TCO calculations. Lower middle panel: Scatterplot of new daily TCO values against traditional TCO. Lower right panel: Scatterplot of daily difference between new and traditional BRDV against number of contributing DS measurements; dark grey markers for data points from months JJASON, red markers from remainder.

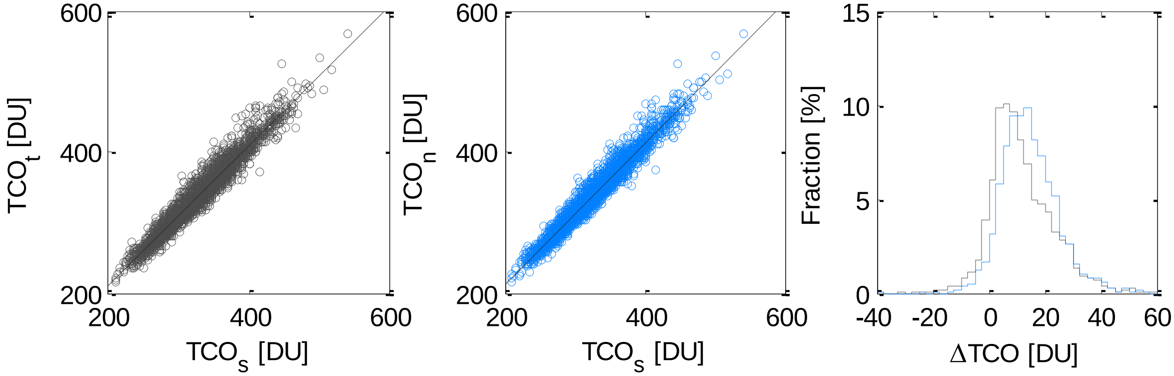

Figure 4. Left panel: Traditional BRDV with traditional DS-ZS choice vs OMI mean overpass. Middle panel: BRDV from described methodology vs OMI mean overpass. Right panel: Histogram of daily differences between traditional (grey) and new (blue) daily TCO calculations vs satellite overpass TCO.