Response to referee's comments

May 2021

1 Anonymous Referee #1

This paper presents two global total ozone data sets based on multiple satellite and ground-based data records. The NIWA-BS data set is a update of a 20 years old data set by Bodeker et al., 2001 and NIWA-BS-filled data set as a new set that is based on NIWA-BS, but has extrapolated values in the areas where no NIWA-BS data exist. These data sets are useful for various applications related to the stratospheric ozone. The paper is well written and can be published after minor revisions.

We would like to thank the reviewer for taking the time to review this paper and for their helpful comments that have led to improvements in the paper.

Major comments:

1. From Figure 7, it appears that the NIWA-BS TCO data set has a negative bias against all validation data sets in the Arctic in winter. This should be investigated further.

At the time that we generated this figure we did investigate this further since the DJF difference in the Arctic did appear anomalous. We found that the TOMS and OMI measurements were generally biased high with respect to the ground-based Dobson and Brewer spectrophotometer measurements over the Arctic in DJF, resulting in a downward correction in the TOMS and OMI measurements at high northern latitudes in this season. Because the corrected TOMS and OMI data then form the basis for the corrections for the other data sets, this induces a downward correction in those data sets. We acknowledge the possibility that the correction required may have been overestimates given that there are only 548 Dobson/Brewer - satellite difference pairs in the Arctic poleward of 60°N. Establishing a better correction for high latitude satellite-based measurements will be a focus for version 3.6 of the NIWA-BS TCO database. Some text has been added to the paper to clarify this.

In addition, how large are the differences for the years with large Arctic depletion (e.g., 1997, 2005, 2011)? Does NIWA-BS TCO capture these events correctly? Perhaps such comparisons can be added to Figure 8.

Recall that the differences are modelled as an offset and trend where the fit coefficients are expanded in Fourier series to capture seasonality and in Legendre polynomials to capture the meridional structure of the differences. Therefore there is no process for capturing year-to-year variability in the differences. This is appropriate as we don't expect satellite-based measurements to show such variability. They are more likely to show some offset and drift.

We have calculated the area weighted Arctic polar cap mean TCO (poleward of 60°N) for DJF and plotted these in Figure 1. In some years it is not possible to calculate a valid DJF mean from the unfilled database due to a paucity of data. The years with low ozone referred to by the reviewer (1997, 2005 and 2011) do indeed show low TCO in both our filled and unfilled databases. Interestingly, 1996 shows a record low; previously, with the unfilled database, a valid value could not be calculated. 2000 also shows very low Arctic polar cap mean TCO with the unfilled value (336 DU) being nearly 25 DU lower than the unfilled value.

2. Validation of the NIWA-BS-filled data set should be done more thoroughly. Why do not you have a day in March or October when global ozone data with no gaps are available. Then keep only data over the areas where data were available on October 31, 1978 and compare the reconstructed field with the real one. This would confirm the uncertainly estimates for the final reconstructed field. The same could be done for November 1, 1978 data to how the algorithm performs in the case of minor gaps.

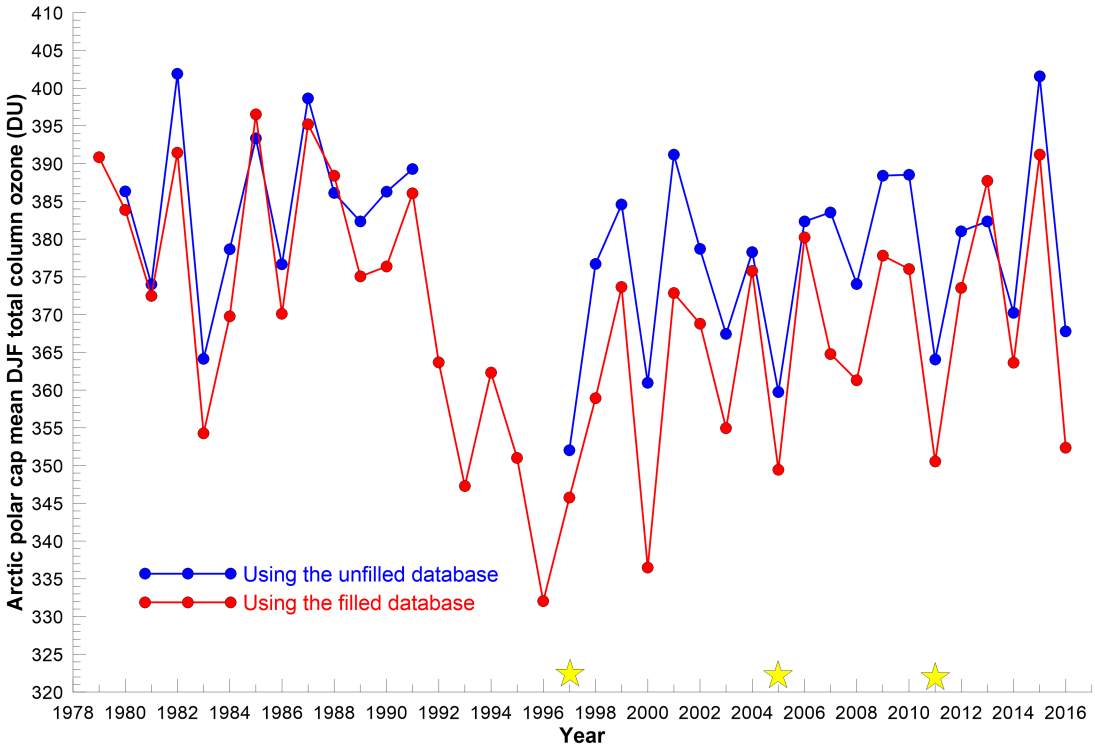

Figure 1: Arctic polar cap area weighted mean TCO for DJF calculated from the filled (red) and unfilled (blue) database. Stars show the year highlighted by the reviewer.

We have done many such tests but did not include these in the paper to keep the paper reasonably short (noting that it already has 18 figures). We take the opportunity to provide some additional validation of the filling method in response to this reviewer.

We have taken the original TCO field on day 80 of 1982 (see Figure 2) and removed swaths of data of with 10, 20, 30, 60 and 120° (see Figure 3).

The data filling algorithm was then applied to the field show in Figure 3 to create the fields shown in Figures 4 and 5.

The differences between the fields shown in Figure 2 and Figure 4 are shown in Figure 6.

The question now is whether the uncertainties shown in Figure 5 are consistent with the differences shown in Figure 6. We use the formalism of Immler et al. (2010) to quantify the extent to which this is true. Specifically we calculate k as:

$$k = \frac{|m_1 - m_2|}{\sqrt{u_1^2 + u_2^2}} \tag{1}$$

where m_1 and m_2 are the original and interpolated data values respectively and u_1 and u_2 are their uncertainties. The field of k values is shown in Figure 7.

There are regions where the uncertainties are under-estimated (k greater than 1.0) and regions where the uncertainties are over-estimated (k smaller than 1.0), but this is to be expected as the uncertainty estimates are, indeed, estimates. The average of all of the k values shown in Figure 7 is 0.892 suggesting that, on average, the uncertainties on the filled values are slightly over-estimated - a value of $k = 1$ denotes that the two values are 'consistent' within their uncertainties (Immler et al., 2010). This is the preferred (conservative) option.

3. There should be some validation of the NIWA-BS-filled data set in the polar night areas. They are the most interesting regions. Some total ozone data, such as moon measurements by Dobsons and Bewers as well as from integrated ozonesonde profiles are available.

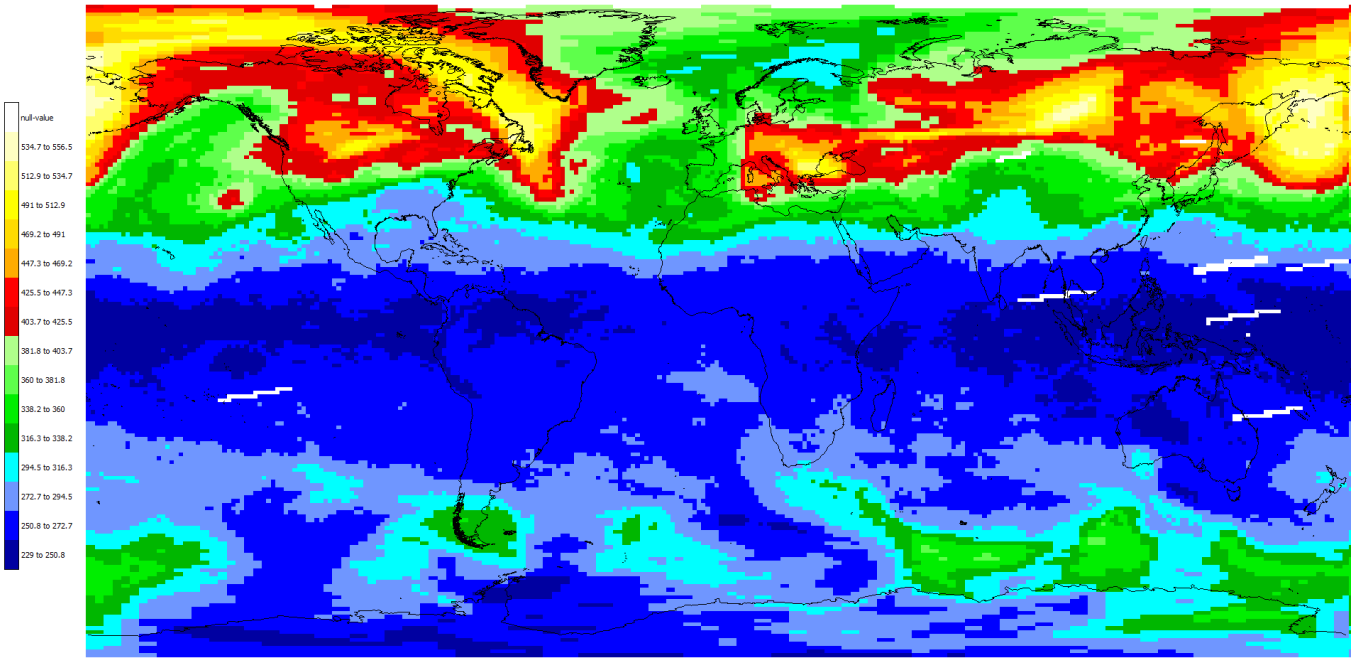

Figure 2: The original TCO field on day 80 of 1982.

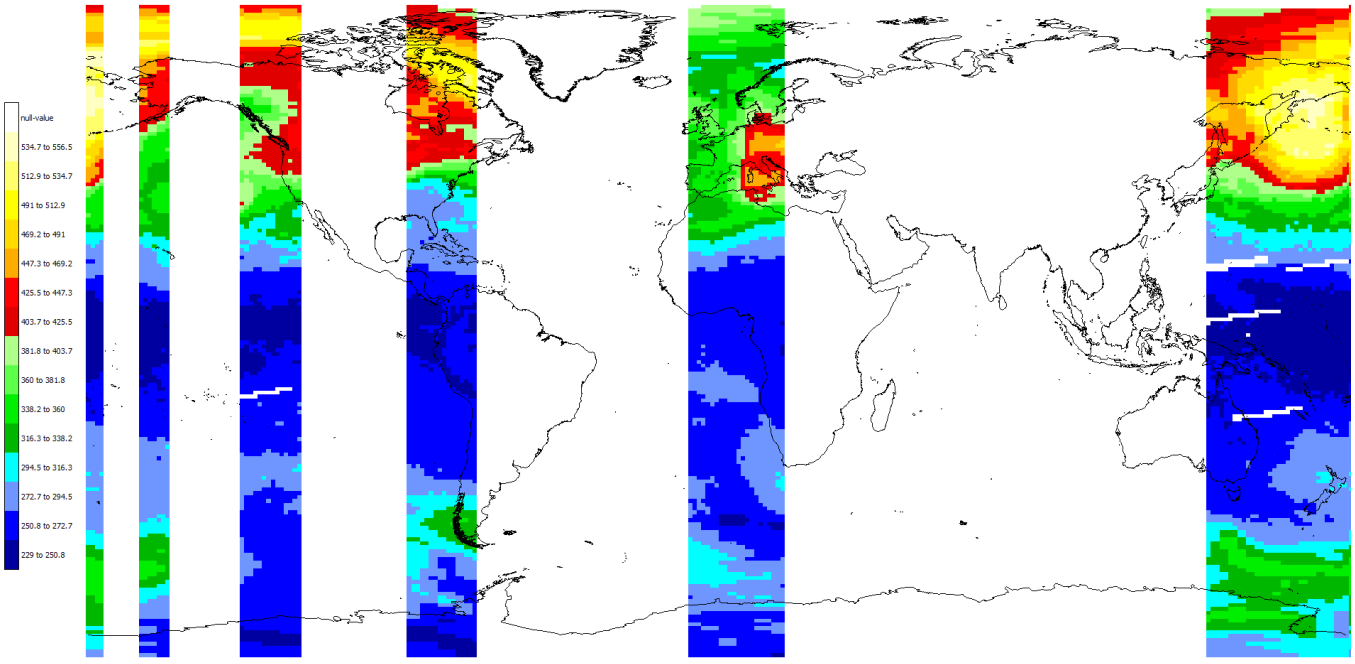

Figure 3: The TCO field on day 80 of 1982 with swaths of width 10, 20, 30, 60 and 120° having been removed.

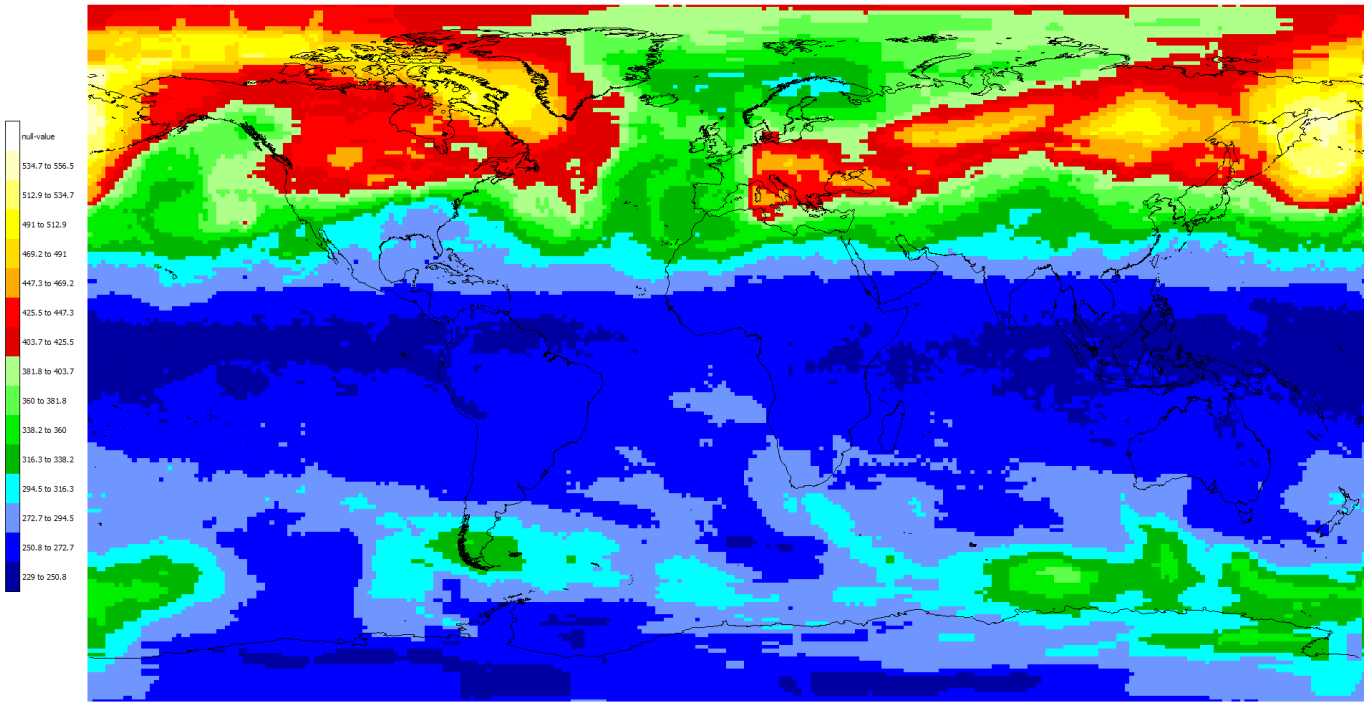

Figure 4: The filled TCO field on day 80 of 1982.

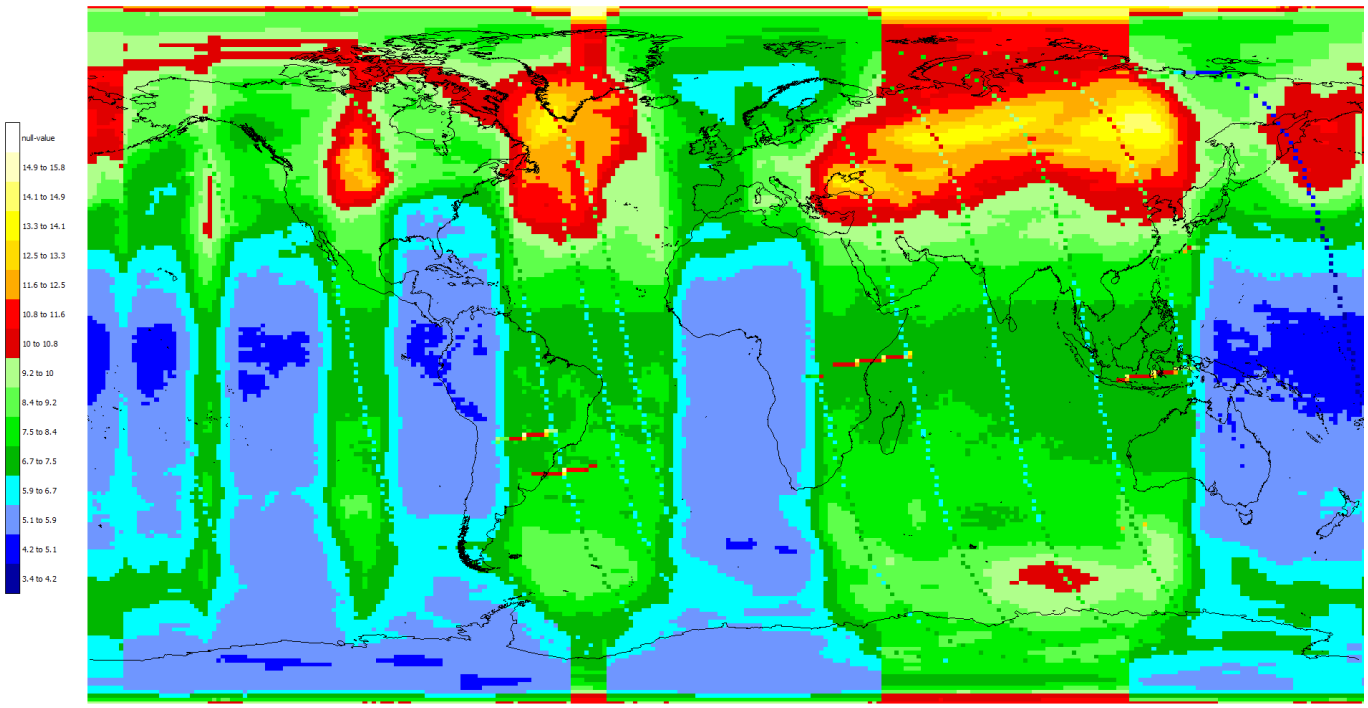

Figure 5: The uncertainties on the filled TCO field for day 80 of 1982.

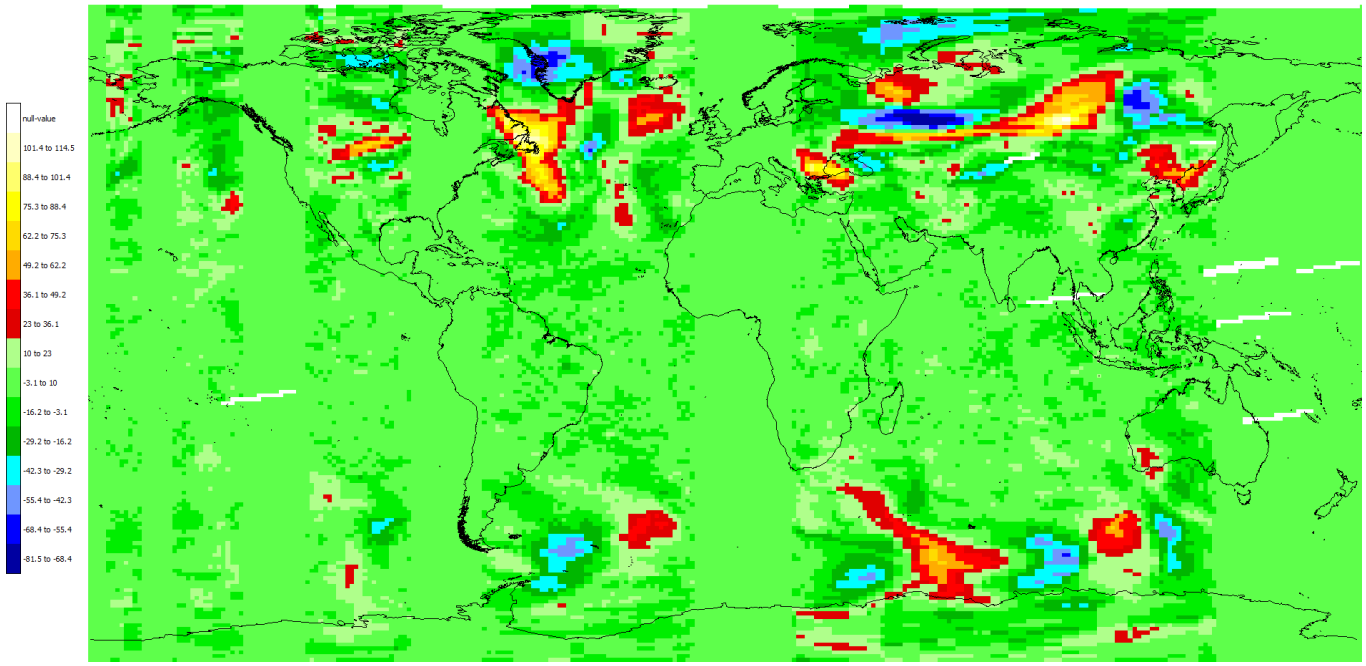

Figure 6: The differences between the original TCO field shown in Figure 2 and the filled field shown in Figure 4.

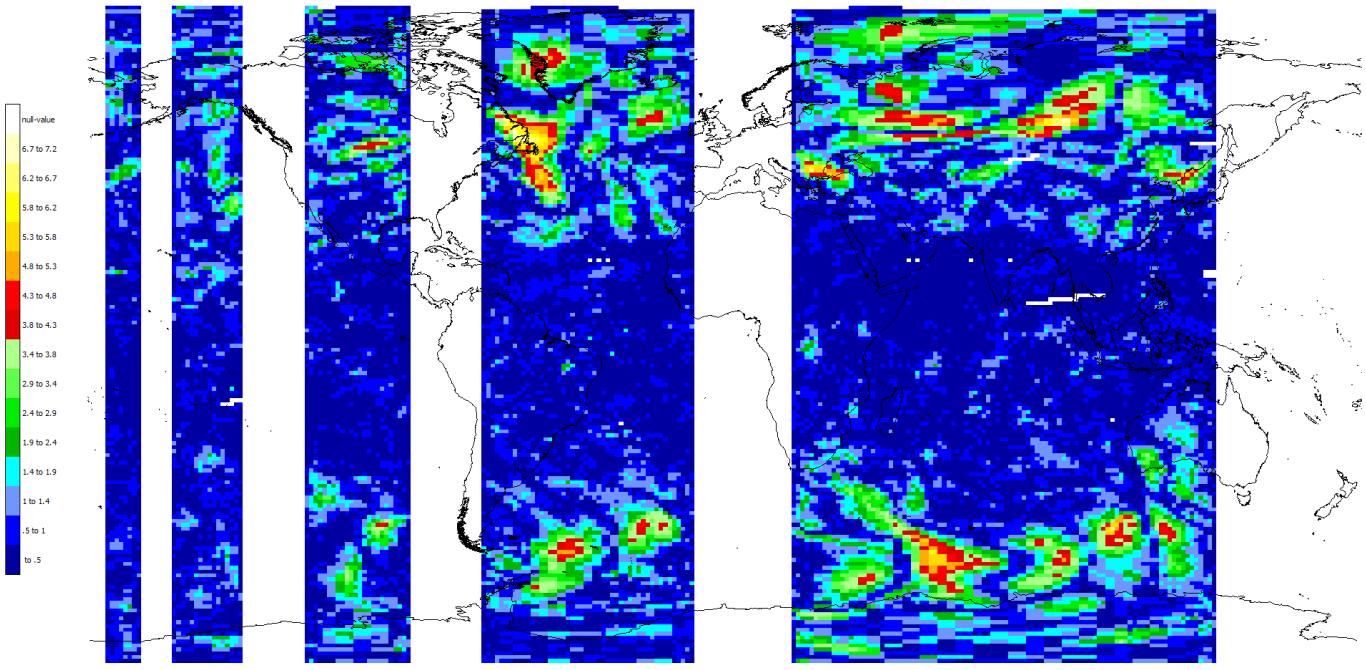

Figure 7: The k values calculated for the filled field. Only k values for the filled values are shown since they are otherwise everyone 0.0 since there $m_1 = m_2$.

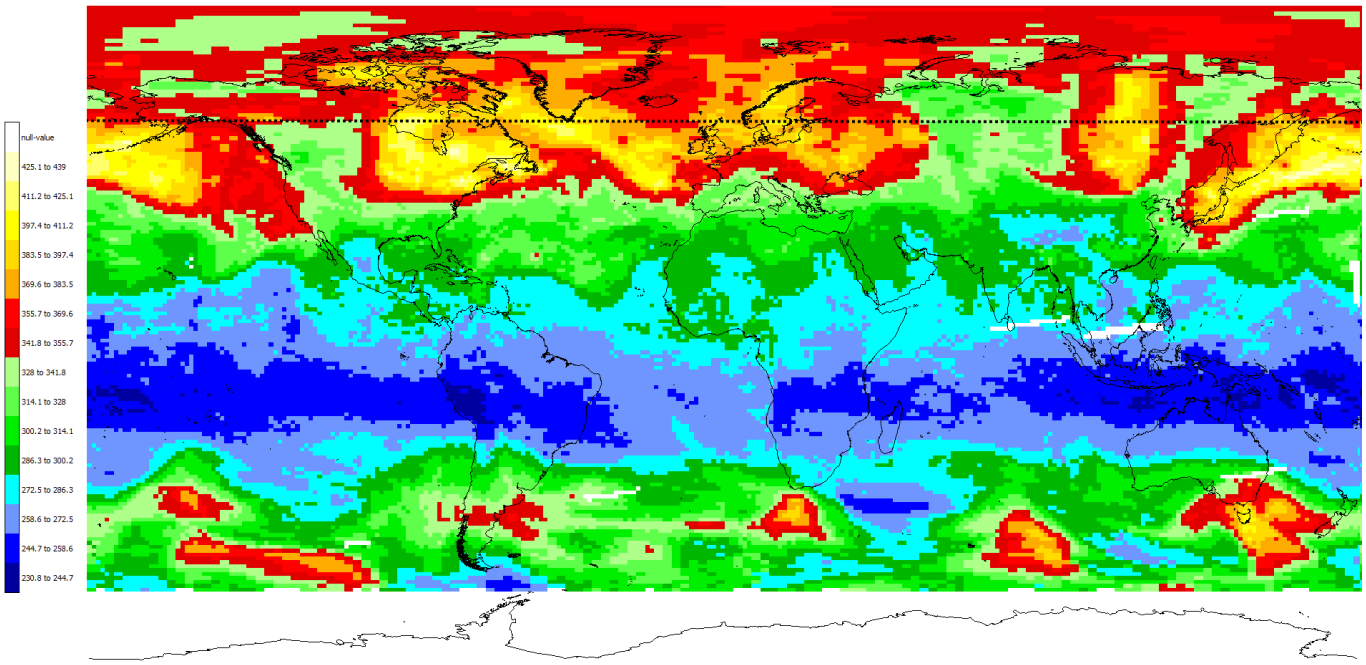

Figure 8: The original TCO field on 21 June 1982. Data north of 60°N were deleted (i.e. poleward of the dashed line).

Rather than trying to source moon measurements from Dobson and Brewer spectrophotometers to conduct the polar night filing validation (noting that these measurements are very sparse both temporally and spatially), we have used a similar method to what was used above to validate the NIWA-BS-filled data set in the polar night areas. Specifically, all data poleward of 60°N where removed over the period 1 June to 15 July from **every** year noting that the machine-learning algorithm also uses data from neighbouring years to learn how TCO is correlated with 550 K PV and tropopause heights. The deleted data were then infilled using the data filling algorithm. We also note that it is necessary to delete data over a long period within the year otherwise the infilling algorithm simply uses data from day $N-1$ and day $N+1$ to fill the missing data. We select for validation 21 June 1982, i.e. i the middle of the period over which the Arctic data were deleted. The original TCO field is shown in Figure 8, the field with data removed poleward of 60°N is not shown (use your imagination), and the field with data filled poleward of 60°N is shown in Figure 9.

In this case the filled TCO values are somewhat overestimated compared to the original data but, on other days, the reverse is true. The question is whether the filled values are consistent with the original data within their uncertainties. Again we calculate the k values as was done above. The map of k values is shown in Figure 10.

The mean of the k values over the Arctic is 1.09 suggesting that the uncertainties calculated on the filled values are realistic in representing the true uncertainty of the filling.

Specific comments:

p.4, Table 1. Do not use tiny.url. They are shorter, but they may not work in a browser. I've checked the links. Some of them do not work, the other require a password.

Thank you for catching this and we have now included direct links and made sure that the data can be found under the links provided. For some data sets, the user will need to create a login and password. The data are freely available but some providers have the requirement to set up an account.

p. 5, l. 2 Why is it assumed that the drift is linear? Have you done any tests?

When we first started this work, around 25 years ago, using TCO measurements from the ground-based Dobson and Brewer spectrophotometer network to correct space-based TCO measurements, we had extensive talks with the

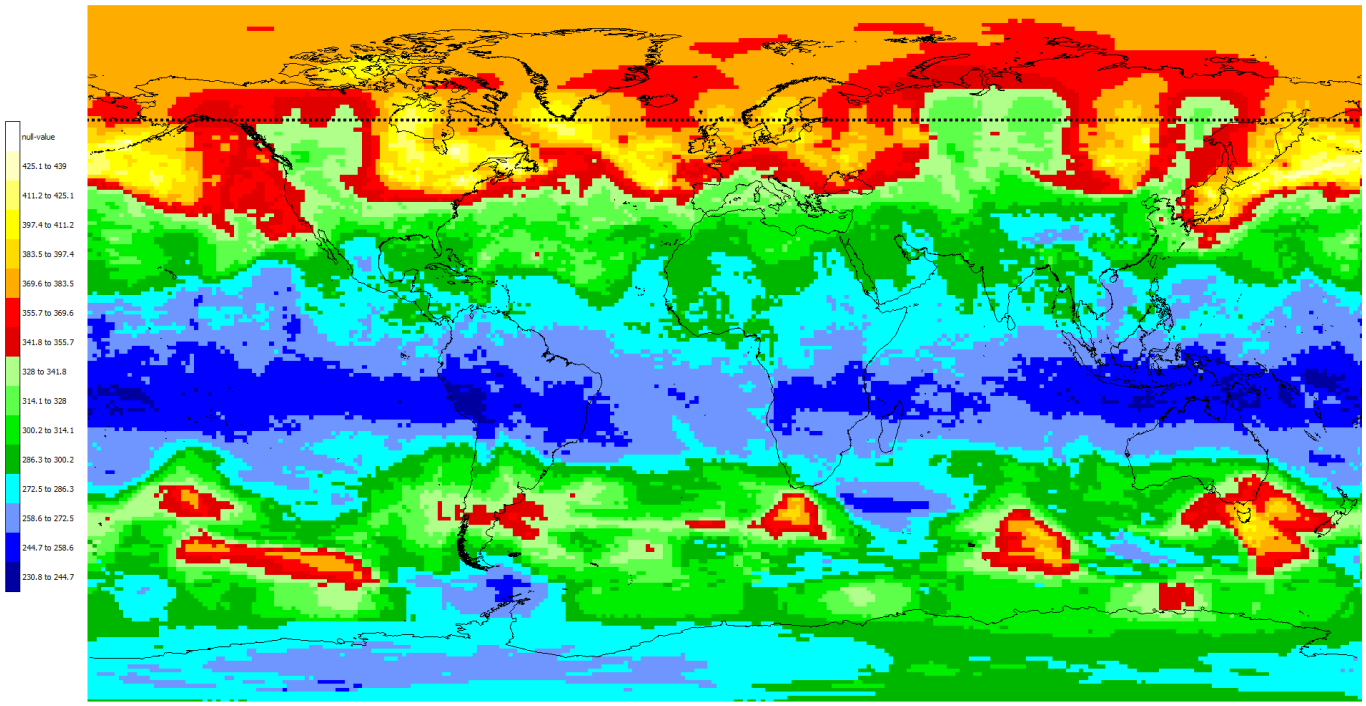

Figure 9: The filled TCO field on 21 June 1982. Data poleward of the dotted line (60°N) were filled.

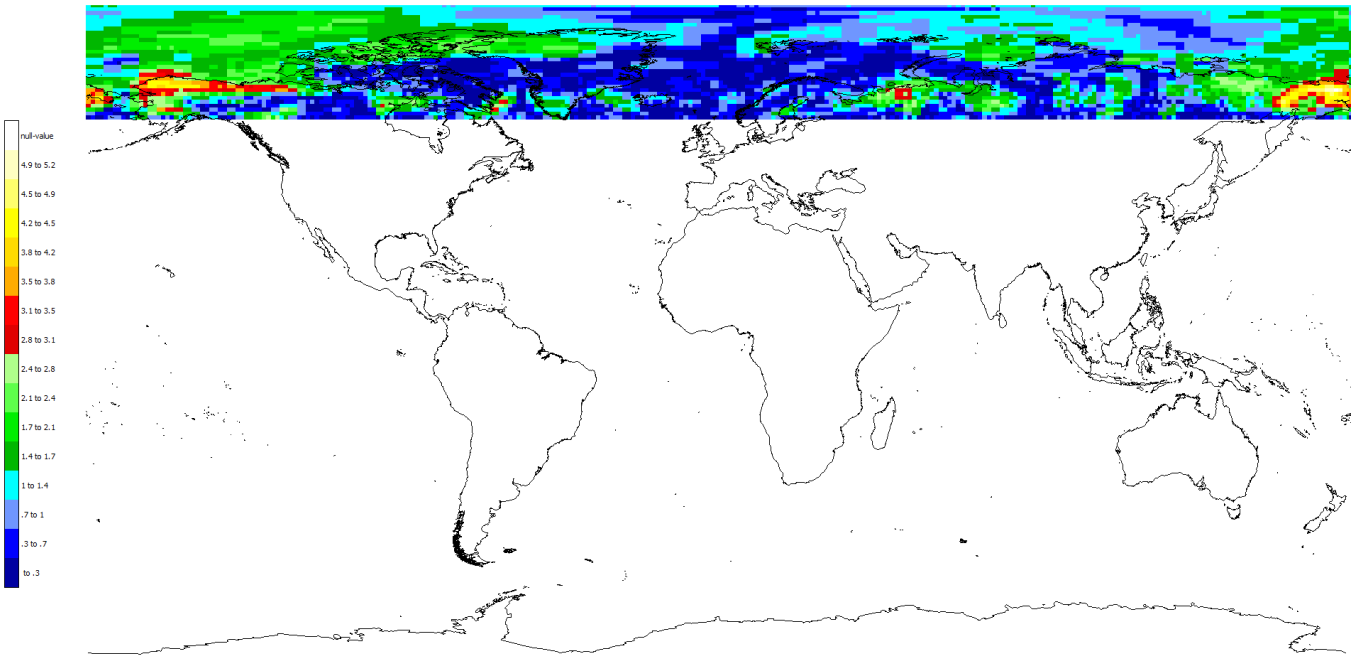

Figure 10: The k values calculated for the Arctic filled data.

Principle Investigators of the space-based instruments (in particular Rich McPeters, PK Bhartia and John Burrows) to discuss what sort of structure in the drift of the satellite-based measurements we might expect to infer from analyses of temporal structure in the differences between the satellite and ground-based measurements. They strongly advised that we consider only linear drift as any other assumption would likely create spurious structures that were not supported by the difference time series (which are noisy, spatially sparse, and where the spatial coverage varies with time). Only where there were well-documented discontinuities, as with the GOME instrument, did we consider structuring the temporal dependence of the differences (i.e. the drift) as anything other than a linear trend. We did some tests nonetheless and found out that their advice was excellent advice - the difference data did not support any assumed structure in the differences other than a linear drift.

p.5, l 15. Some GB stations (e.g. Mauna Loa) have a bias with satellite data due to high elevation that is not properly accounted by large satellite pixels.

Yes, for the few sites that are at the top of very high and pointy mountains, a good portion of the tropospheric partial ozone column will be lost to the ground-based measurement but will be seen by the satellite (even in overpass mode) which will likely have a much bigger footprint than the peak of the mountain. The corrections that we derive for the satellite-based measurements are largely impervious to this effect since the corrections never rely on any single site. Rather, we fit a 2D surface (as a function of latitude and time, expanded in Legendre polynomials and Fourier series respectively) to the differences calculated at individual sites. The 'stiffness' of this surface, defined by a limited Legendre polynomial expansion, makes the statistically modelled differences field insensitive to such single site effects. That said, in future version of the creation of this data-base, we will exclude all sites from the difference analyses whose altitudes are above 1 km. We thank the reviewer for bringing this to our attention.

p.5, l 15. Dobson and Brewer instrument have different dependence of stratospheric temperature. This introduces a seasonal difference that could be as high as 2%. Ideally, Dobson data should be adjusted using effective stratospheric temperature.

We are aware of this effect from papers such as Bernhard et al. (2005) and Balis et al. (2007). We are also aware that some recent reprocessing of the Dobson spectrophotometer data has corrected for this effect. Rather than trying to figure out where this correction has and hasn't been applied, and applying it where it hasn't, we have trusted that the TCO data records provided by the World Ozone and UV Data Centre have been processed to be as accurate as possible. As such, we have applied no further corrections to the Dobson spectrophotometer data.

p.5, l 15 What Dobson and Brewer data were used – all data? DS only?

It was stated in the paper that we use direct-sun measurements only. We add that we also used measurements only made with the AD wavelength pair (both ordinary setting and focused images) for the Dobson spectrophotometers, and all data from the Brewer spectrometers.

What about SAOZ data? They could be very useful at high latitudes.

We assessed the volume of SAOZ data that would have been available for such corrections and considered it to be too small to warrant the additional effort to include these data when inferring the satellite-based instrument biases. It was for the same reason that we didn't use DOAS retrievals, microwave radiometer retrievals, or integrated ozonesonde or lidar ozone profiles. But, yes, the availability of SAOZ TCO measurements at high latitudes could be helpful. We will reassess this data source in future constructions of this database. We thank the reviewer for raising this idea.

P. 16-18. Fig 9-11 are not very informative. Perhaps they should be in a supplement

These figures show the spatial and temporal complexity of the structure between our TCO database and the other validation data sets. As such, we believe it is valuable to keep these figures. These figures show in much greater detail the variability in the TCO differences in each month and latitude band; information that is lost when calculating means as in Figure 7. The other two referees felt that these figures had value.

p. 19, Fig 12. It may be better to show some interesting periods/events rather than "twelve selected months/years"

We decided to follow an objective method for which of the many monthly difference fields between the NIWA-BS and longitudinally resolved ESA CCI database to show. We felt that selectively picking interesting periods/events may only create increased opportunity of perceived selection bias - what may be 'interesting' to us may be completely uninteresting to some other reader who may have, e.g., an interest in the bias in some particular month over some other. We decided, therefore, to show one field for each month of the year and to move sequentially through the years for which difference fields were available. We have updated the figure accordingly.

p. 27, l. 24. Please justify why 2000 was used as the trend turning point.

We have added two sentences to the paper to explain this, i.e. '1999/2000 was prescribed as the trend transition year as this is approximately when stratospheric chlorine and bromine loading peaked (Newman et al., 2007). We also wanted to ensure that the first trend period included data from the late 1990s as there was a greater likelihood of missing data from 1994 to 1998 and we wanted to avoid end-effect-biasing in the calculation of the trends. That said, the conclusions drawn below regarding changes in trend were found to be largely insensitive to the selection of the transition year within 2 years of the selected transition year.'
We thank the reviewer for prompting us to add this detail. It was clearly an omission in the first version of the paper.

2 Anonymous Referee #2

The paper presents the datasets of total ozone column (TCO), which are created using the data from several satellite measurements. One dataset is an improved version (v3.4) of NIWA-BS TCO dataset, while another is gap-free BS-filled TCO database. The paper describes the methods used in the construction of the datasets, and some evaluations of dataset, including evaluation of ozone trends. These datasets are valuable contributions to the ozone research. The paper is well written. Several minor comments and suggestions for paper improvement are below.

MAIN COMMENTS

1. The differences between individual datasets are evaluated using zonal mean values. Does this approach work also in presence of polar vortex? Please add a discussion or evaluation.

Yes, other than for the four TOMS and OMI data sets where the biases are calculated from comparisons between satellite-based and ground-based measurements at individual sites (but them modelled without zonal structure), corrections for the other data sets are determined through comparisons of zonal means against the corrected TOMS and OMI data. If the zonal sampling of a satellite was biased, e.g. if the satellite only made measurements between say 0°E and 90°E, then its zonal mean could be very biased with respect to the zonal mean from an unbiased satellite-based instrument, especially in the presence of zonal asymmetries as occurs when the polar vortex is pushed off the pole, e.g. by a large planetary wave of zonal wavenumber 1, as is pointed out by the reviewer. But this is seldom, if ever, the case, i.e. the TCO measurements from the satellites we have used are zonally unbiased and the comparisons do not suffer from such effects. Cognisant that this may have not been the case for the more spatially sparse SBUV measurements, we did include an additional source of uncertainty in the calculation of SBUV zonal means, as discussed on pages 9 and 10 of the original paper.
We have added material towards the top of Section 4 to this effect, i.e. 'There could be a danger here that in the case of biased longitudinal sampling by one satellite compared to another, that the zonal means would be biased but without these differences arising from any intrinsic biases between the satellite-based measurements. Only the SBUV measurements were sparse and corrections for this potential sampling bias were derived as discussed below.'

2. P.18: "For some applications, there is a need for gap-free TCO fields". Please indicate these applications.

We have now expanded this sentence to say 'For some applications, there is a need for gap-free TCO fields, e.g., TCO fields for validating chemistry-climate models which generate TCO fields over the entire globe for each day of the year.'

3. Conservative filling algorithm (Section 9.1). In general, it is easy and advantageous to demonstrate the quality of the filling procedure with artificially created missing data: from a full field some data can be masked and the filling algorithm applied. Then the quality of filling and the quality of uncertainty estimates can be directly evaluated

using the true and reconstructed data.

This is a good idea and the same suggestion was made by Referee 1. The results of exactly that sort of evaluation (an additional 10 figures), are provided in the response to Referee 1. We decided not to include all of this additional validation of the filling method in the paper as (1) the paper already has 18 figures and adding another 10 would have been untenable, and (2) because the response to the referee comments accompany the paper, we felt that readers who were interested in a deeper validation of the filling technique could refer to our response to Referee 1.

For the conservative filling, using the data from previous of following day is a dangerous operation, in my opinion. The air masses are moved, thus such interpolation can result in significant errors, especially in regions of high ozone gradients.

We never just use the value from the previous or following day. As stated in the paper, we linearly interpolate between values on day $N+1$ and day $N-1$ to estimate a value for day N. We have found this to be very robust and as can be validated by inspecting a TCO time series for any location on the global. Given the very long photochemical lifetime of TCO, the time series tend to be rather smooth suggesting that linear interpolation between neighbouring days is an acceptable way to estimate missing data. The assumption that is made is not that ozone is unchanged from one day to the next, but that TCO changes approximately linearly from one day to the next. Essentially this is equivalent to a CFL (Courant–Friedrichs–Lewy) condition for TCO.

The calculation of uncertainties is not described in detail. In particular, it is unclear how the distance from available measurements is taken into account, and this is not seen in Figure 13.

Yes, we have not described the calculation of the uncertainties in the longitudinal linear interpolation in great detail. We were trying to keep the paper as succinct as possible. Calculating the uncertainties on such longitudinally interpolated values is non-trivial. The challenges of estimating the uncertainties on such interpolations are described nicely in **?**. We have used a rather simple method for estimating the uncertainty on the longitudinal interpolation, as follows:

$$\begin{aligned} \alpha &= (\phi - \phi_1)(\phi_2 - \phi_1) \\ T_1 &= (1-\alpha) \times \sigma_1^2 \\ T_2 &= \alpha \times \sigma_2^2 \\ \sigma_{interp} &= \sqrt{T_1 + T_2} \end{aligned} \tag{2}$$

where ϕ is the longitude at which the interpolation is being performed, ϕ_1 is the longitude to the west where there is a non-null value, ϕ_2 is the longitude to the east where there is a non-null value, σ_1 is the uncertainty on the non-null value to the west and σ_2 is the uncertainty on the non-null value to the east.
The reason that this inflation of uncertainties in conducting the longitudinal interpolation is not apparent in Figure 13, is that much of the interpolation in Figure 13 is through interpolation between day $N-1$ and day $N+1$ which takes precedent over the longitudinal interpolation.

4. Machine learning estimated ozone: It is important to assess the quality of this approach in unexpected ozone conditions, for example Arctic ozone hole in 2011 and 2020. For example, data from 2011 can be excluded from the training dataset, and then tried to reproduce. In general, demonstration of ozone hole evolution in the created datasets would be a very interesting and valuable addition to the paper.

We agree and this has been done in some detail in response to Reviewer 1. We have not explicitly done this for 2011 and 2020. In suggesting that some validation can be achieved by withholding, e.g., 2011, from the data set, we believe that the reviewer is seeing this as a neural network-based solution. It is not. Rather, it is a regression approach where the regression model establishes a local (in space and time, where neighbouring years are included in the time window) statistical dependence of TCO on local tropopause height and local potential vorticity at the 550 K surface. We could train the regression model to generate TCO fields for all of 2011 when all TCO fields from 2011 are excluded, but this would not be an appropriate validation since the regression model has been constructed to assume that there will be TCO fields within a few days of the target field which is always the case. We have not done as the reviewer has suggested for 2020 since our database terminates in 2016.

5. Section 10. More details on the regression model would be useful. In particular (a) If possible, repeat the

sources of proxies (P.27, L.28) instead of reference to Bodeker et al. (2001),

We did not include details of the El Niño Southern Oscillation (ENSO), solar cycle, Mt. Pinatubo and El Chichón basis functions in this paper as they have been published in Bodeker et al. 2001b and Bodeker et al. 1998, and both papers are publicly available. Including the additional information here is unnecessarily verbose and would lengthen the already long paper even further.

(b) Please add a note on performance of this global fit when data from some months are missing (in polar night conditions).

It is not clear to us what the reviewer is referring to here as the fit is not global, i.e., equation (18) is applied completely independently at each latitude and longitude. As to the challenges of obtaining trends during the polar night, while the Fourier expansion in the regression model fit coefficient (accounting for trend structure at annual, 6-monthly, and 4-monthly periods) is somewhat robust against data in missing months, there is a danger of over-fitting at high latitudes in winter and obtaining unreliable trends in regions of polar darkness. This is why we exclude trends in these regions in Figures 16 and 17.

Please add the figure with TCO trends after 2000 (in addition to change trends). This will allow direct comparison with other studies.

Yes, this is a good idea and we have added this figure as the new Figure 18.

SPECIFIC COMMENTS

1) Please write direct links to the datasets, not via tinyurl.com, which do not work properly.

Thank you for catching this and we have now included direct links and ensured that the data can be found under the links provided.

2) P.4: A map showing locations of Brewer & Dobson stations would be useful.

We have decided not to do this for two reasons:

1. It would add yet another figure to the paper.
2. As a nice interactive map of all WOUDC sites is provided at `https://woudc.org/data/stations/index.php`.

We felt that an additional figure of all the measurement sites is unwarranted and any reader will find the information (easily available) on the well maintained and updated WOUDC website. We have added a pointer to the this URL in the paper towards the beginning of Section 3.

2)Also a statement on compatibility/similarity of Dobson and Brewer data and their quality would be useful.

We have added the following sentence to the paper 'We assumed in all cases that the Dobson and Brewer spectrophotometer data submitted to the WOUDC were the highest quality data available, that all possible corrections to improve the quality of the data had been made prior to submission, and that the measurements from these two networks were unbiased with respect to each other. Assumed uncertainties on the measurements from these two networks are presented below'. We felt in unnecessary to provide a more detailed discussion of the uniformity of the data either within or between these two measurement networks as many other papers have conducted such analyses, e.g. Bojkov et al. (1986), Bojkov et al. (1990), Zerefos et al. (1992), Bokjov et al. (1995).

3) P.4, L1: Please clarify what you mean by "a higher quality data set": a higher spatial resolution?

We have replaced this sentence with 'While TOMS and OMI data are provided in gridded data files, the original overpass data, unlike the gridded data set where several measurements from different times through the day might be averaged within a grid cell, provide location-specific measurements that are more suitable for comparison with the ground-based measurement networks, i.e. the overpass measurements are specific to a latitude, longitude and date/time.' to be more explicit in what we had previously meant when we had only said 'a higher quality data set'.

4) P.5, around Eq.(2): Since you use integer numbers for f, please indicate also the units of t. It is also worth to note that the choice Nf,b=0 corresponds to the assumption of a constant drift.

We have added just after equation (1) that time is measured in decimal years. The choice of $N_{F,\beta} = 0$ is not that the drift is constant but that the drift is seasonally independent. A sentence has been added to the paper to clarify this.

5) P.7, Fig.3 caption: "Regions shaded in grey" –should be "in black"?

Thank you and we have corrected this error.

6) Table 2. Please indicate that all uncertainties presented in the table are averaged/typical values, which can be different from the uncertainties for a particular location and time.

As suggested by the reviewer, we have amended the table caption to 'Typical uncertainties on the source data sets as reported in the references provided and as used in the construction of the TCO databases. Note that the uncertainty on any particular measurement may differ from the typical values quoted in this table'. We have also removed the links from Table 2 and included the links as references.

7) P.9 L 2: should be sigma_Delta (big Delta)?

The reviewer is correct and we have changed the small delta to a big delta.

8) P.9, Eq.(6), sigma_i should be squared

Thank you for spotting this mistake, which we have now corrected.

9) P.10, L.5. Some data are on 1x1 deg grid. Please clarify how the re-gridding is done.

Where source data were provided at $1^\circ \times 1^\circ$ resolution, bilinear interpolation was used to resample the data to $1.25^\circ \times 1^\circ$ resolution. A sentence to this effect has been added early in Section 2.

10) Fig.7. It would be useful to indicate also standard deviation of differences.

We agree and have made a new version of Figure 7.

11) Figures 9 -11: please use smaller color limits.

If by 'use smaller color limits' the reviewer means that we should reduce the colour range on each plot, we prefer not to for two reasons: (1) We would like to capture the extreme values and (2) we would like to use the same color range on each figure to make them comparable.

12) Eq.9, LHS: comma instead of "-"

This has been corrected.

13) P.23, lines 22-26: The description is not clear, perhaps, an illustration (can be put in the Supplement) would be useful.

This paper is not accompanying by a Supplement and we were reluctant to now add a Supplement just to accommodate this single additional figure. Rather than adding a figure to the paper, we have described in greater detail how the A_{proxy} values are calculated.

3 Anonymous Referee #3

This study improved the long-term total column ozone data record by extending the relevant satellite data input and statistical methods, as well as adding the uncertainty level accompany the released datasets. As an important data source for climate change study, the update of long-term ozone distribution would be helpful for the community to investigate ozone-related topics and analyses. The manuscript could be improved by addressing the following issues:

1) I would suggest to mark the added satellite data (Figure 1) in a way to show the difference with previous data records.

We have now done this and have added explanatory text to the figure caption. Newly added satellite are shown with an additional dashed line.

2) P6 L9, what's the 'additional basis function' in particular? Is that simply' set to zero prior to 22 June 2003 and to 1 thereafter'?

Yes, as stated in the manuscript. This basis function then allows for an additional offset across the 22 June 2003 transition.

3) By spanning the temporal coverage of each of those non-TOMS/OMI data sets, would this implementation introduce for data sets not covering the time span shown in Figure 3?

We do not understand the referee's comment. As shown in Figure 1, the TOMS+OMI record spans the full period of the data set except for a gap of approximately 1 year in 1995. As such, the corrected (biases and offsets removed), TOMS+OMI data set provides a valid standard against which all other satellites can be compared.

4) P8 L17, what's the setting for atmosphere when simulating the uncertainties using Monte Carlo? Personally, I would like to know what to come up a way to evaluate the uncertainty prorogated from the data sources to the analysis.

We don't understand what the referee is referring to by "what's the setting for atmosphere". Essentially we need to calculate uncertainties on the 2D manifold (latitude and time) fitted to the difference pairs and propagate those uncertainties through to the final data product, i.e. the derived corrections are themselves uncertain and when applying those corrections an additional uncertainty is introduced which must be included into the uncertainty budget. We do this by generating different, but statistically equivalent, versions of the difference data set by Monte Carlo sampling the uncertainties on the differences and adding those to the original differences. We create 100 such manifolds. The mean and standard deviation of the 100 manifolds provides the final difference field and its uncertainty.

5) P10, L5, Would it be possible to further increase the resolution of this product, like length of each side per pixel of 1° or even higher?

No. Essentially we are constrained by the coarsest resolution of the source data sets which is $1.25° \times 1.0°$ for the TOMS data.

6) The polar night shown in Figure 4 is shaded with a different color.

True. This is stated in the figure caption.

7) I noticed that the latest three years 2013-2016 is not covered in the validation (Table 3), why?

Thanks for catching this and we have made the necessary corrections to Table 3.

8) Figure 7, The uncertainties seems to be largest around latitudes of 60°, please discuss the reason.

We have not diagnosed why the uncertainties on the NIWA-BS TCO values maximize around 60° latitude. That's simply what emerged from tracing the various sources of uncertainty through to the final data product.

In diagram of Dec/Jan/Feb, why the lines go beyond -20 around 60-90°?

A detailed discussion of the large differences between the NIWA-BS TCO database and the other validation databases is given in response to comment 1 of Referee 1. We refer the referee to that response.

9) Is that the areas close to polar night regions are expected to have larger uncertainties?

While we are not entirely sure what the referee is referring to here, we can state that because TCO values tend to be higher at higher latitudes, and that we have assumed fixed percentage uncertainties on many of the source data sets, that the absolute uncertainties tend to maximize at higher latitudes. For the filled version of the database it is certainly true that the uncertainties maximize over the polar night regions since this is where the ML-based filling occurs and these can be accompanied by large uncertainties.

10) When using nearest neighbour interpolation to fill missing values, will you test the area of the gap, it might be misleading for filling large areas of missing values.

Yes, as stated in the paper, for the nearest neighbour interpolation, only single cells neighboured by non-null cells are filled.

11) Filling the gap is essentially useful for applications, will the product indicate the regions that is filled with machine learning, while adding uncertainties for these regions might be challenging.

There are two ways in which users of the database can detect where filling has occurred:

1. Examine both the unfilled and filled database and see where the latter has values and the former does not.
2. Examine the uncertainties on the filled TCO fields. Uncertainties tend to be much higher where values have been filled.

We agree with the reviewer that deriving uncertainties for the filled values is challenging. A validation of the uncertainties on ML-filled regions is provided in response to comment 3 of Referee 1. We refer this referee to that response.

References

Balis, D., Kroon, M., Koukouli, M.E., Brinksma, E.J., Labow, G., Veefkind, J.P., and McPeters, R.D.: Validation of Ozone Monitoring Instrument total ozone column measurements using Brewer and Dobson spectrophotometer groundbased observations, J. Geophys. Res., 112, D24S46, doi:10.1029/2007JD008796, 2007

Bernhard, G., Evans, R.D., Labow, G.J., and Oltmans, S.J.: Bias in Dobson total ozone measurements at high latitudes due to approximations in calculations of ozone absorption coefficients and air mass, J. Geophys. Res., 110, D10305, doi:10.1029/2004JD005559, 2005

Bojkov, R.D., Bishop, L., and Fioletov, V.E.: Total ozone trends from quality-controlled ground-based data (1964-1994), J. Geophys. Res., 100D12, 25867-25876, 1995

Bojkov, R., Bishop, L., Hill, W.J., Reinsel, G.C., and Tiao, G.C.: A statistical trend analysis of revised Dobson total ozone data over the Northern Hemisphere, J. Geophys. Res., 95, D7, 9785-9807, 1990

Bojkov, R.D.: The 1979-1985 ozone decline in the Antarctic as relected in ground based observations, Geophys. Res. Lett., 13, 12, 1236-1239, 1986

Fassò, A., Sommer, M., and von Rohden, C.: Interpolation uncertainty of atmospheric temperature radiosoundings, Atmospheric Measurement Techniques, doi:10.5194/amt-2020-161, 2020

Immler, F.J., Dykema, J., Gardiner, T., Whiteman, D.N., Thorne, P.W., and Vömel, H.: Reference Quality Upper-Air Measurements: guidance for developing GRUAN data products, Atmospheric Measurement Techniques, 3, 1217–1231, doi:10.5194/amt-3-1217-2010, 2010.

Zerefos, C.S., Bais, A.F., Ziomas, I.C., and Bojkov, R.D.: On the relative importance of quasi-biennial oscillation and El Nino southern oscillation in the revised Dobson total ozone records, J. Geophys. Res., 97, D9, 10135-10144, 1992

A Global Total Column Ozone Climate Data Record

Greg E. Bodeker[1,2], Jan Nitzbon[1,3], Jordis S. Tradowsky[1], Stefanie Kremser[1], Alexander Schwertheim[1], and Jared Lewis[1]

[1]Bodeker Scientific, 42 Russell Street, Alexandra, 9320, New Zealand
[2]School of Geography, Environment and Earth Sciences, Victoria University of Wellington, New Zealand
[3]Permafrost Research Section, Alfred Wegener Institute Helmholtz Centre for Polar and Marine Research, Potsdam, Germany

Correspondence: G. E. Bodeker
(greg@bodekerscientific.com)

Abstract. Total column ozone (TCO) data from multiple satellite-based instruments have been combined to create a single near-global daily time series of ozone fields at 1.25° longitude by 1° latitude spanning the period 31 October 1978 to 31 December 2016. Comparisons against TCO measurements from the ground-based Dobson and Brewer spectrophotometer networks are used to remove offsets and drifts between the ground-based measurements and a subset of the satellite-based measurements. The corrected subset is then used as a basis for homogenising the remaining data sets. The construction of this database improves on earlier versions of the database maintained first by the National Institute of Water and Atmospheric Research (NIWA) and now by Bodeker Scientific (BS), referred to as the NIWA-BS TCO database. The intention is that the NIWA-BS TCO database serves as a climate data record for TCO and, to this end, the requirements for constructing climate data records, as detailed by GCOS (the Global Climate Observing System) have been followed as closely as possible.

This new version includes a wider range of satellite-based instruments, uses updated sources of satellite data, extends the period covered, uses improved statistical methods to model the difference fields when homogenising the data sets, and, perhaps most importantly, robustly tracks uncertainties from the source data sets through to the final climate data record which is now accompanied by associated uncertainty fields. Furthermore, a gap-free TCO database (referred to as the BS-filled TCO database) has been created and is documented in this paper. The utility of the NIWA-BS TCO database is demonstrated through an analysis of ozone trends from November 1978 to December 2016. Both databases are freely available for non-commercial purposes: the doi for the NIWA-BS TCO database is 10.5281/zenodo.1346424 (Bodeker et al., 2018) and is available from https://zenodo.org/record/1346424. The doi for the BS-filled TCO database is 10.5281/zenodo.3908787 (Bodeker et al., 2020) and is available from https://zenodo.org/record/3908787. In addition, both data sets are available from http://www.bodekerscientific.com/data/total-column-ozone.

1 Introduction

Total column ozone (TCO) has been identified as one of 50 essential climate variables (ECVs) by GCOS (Global Climate Observing System; GCOS-138 (2010); Bojinski et al. (2014)). Climate data records of ECVs serve a variety of purposes, e.g. climate data records of TCO are required to (1) assess the impacts of changes in ozone on radiative forcing of the climate

system, (2) assess the effectiveness of the Montreal Protocol for the protection of the ozone layer, and (3) determine the contribution of ozone changes to observed long-term trends in surface UV radiation. This paper presents an update of a database which has been used in many previous studies (e.g. Bodeker et al., 2001a, b, 2005; Müller et al., 2008). The database was first developed by NIWA (the New Zealand National Institute of Water and Atmospheric Research) and, in the last decade, has been maintained and updated by Bodeker Scientific (BS). The non-filled database is hereafter referred to as the NIWA-BS TCO database and the filled database is referred to as BS-filled TCO database. The version 3.4 (V3.4) database reported on here extends from 31 October 1978 to 31 December 2016. In constructing this database, the guidelines for generating climate data records of ECVs detailed in GCOS-143 (2010) have been adhered to.

Improvements over earlier version of the database implemented in V3.4 include:

- New and updated sources of satellite-based TCO measurements are used, viz. data from NPP-OMPS (National Polar-orbiting Partnership-Ozone Mapping and Profiler Suite), GOME-2 (Global Ozone Monitoring Experiment-2) and SCIAMACHY (Scanning Imaging Absorption Spectrometer for Atmospheric CHartographY) are now included in the combined data set. Various updates to previously used data sets are detailed in Section 2.

- Improved statistical methods are used to model the difference fields between data sets; zonal means of the difference fields are modelled using Legendre expansions which comprise the meridional component of a spherical harmonic expansion which is best suited for statistically describing a field on a sphere (see Section 3 for more information).

- Measurement uncertainties on the source data sets, and the corrections applied to those data sets, have been collated and are propagated through to the combined ozone data set so that, for the first time, this data set is now provided with uncertainty estimates for each datum.

- Furthermore, the gap-free BS-filled TCO database has been generated (see Section 9) using a machine-learning (ML) algorithm that is trained to capture the broad-scale morphology of the TCO field which extends to regions where measurements are missing. The ML algorithm is based on regression of available data against NCEP (National Centers for Environmental Prediction) CFSR (Climate Forecast System Reanalysis) reanalysis tropopause height fields and against potential vorticity (PV) fields on the 550 K surface.

2 Source data

The various satellite-based TCO data sets used to create the version 3.4 NIWA-BS TCO database are summarized in Table 1. Where source data were provided at $1° \times 1°$ resolution, bilinear interpolation was used to resample the data to $1.25° \times 1°$ resolution. The time periods covered by the satellite data sets are shown graphically in Fig. 1. In addition to the information presented in Table 1:

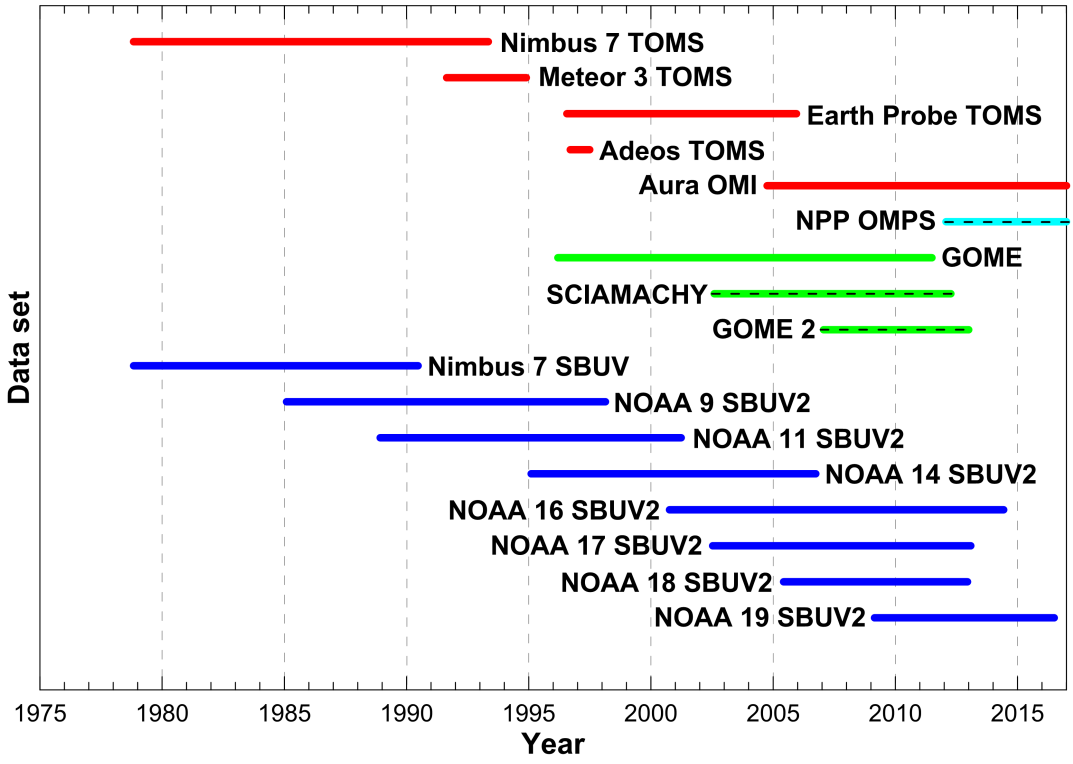

Figure 1. A graphical representation of the different satellite-based data sets used in this study and their periods of coverage. Satellites marked with a dashed line indicate newly added source data for this version of the database.

- The four TOMS (Total Ozone Mapping Spectrometer) data sets (Adeos, Earth Probe, Meteor-3 and Nimbus-7) all use the TOMS retrieval algorithm with Adeos using the version 7 algorithm and the remaining three using the version 8 algorithm.
- An algorithm similar to that used in the version 8 TOMS retrieval is used to conduct the version 8.6 retrievals of the SBUV (Solar Backscatter UltraViolet instrument) data (McPeters et al., 2013). Sparse gridded files were generated from the SBUV TCO measurements made at discrete locations as input to the process that creates the combined database.
- The high resolution and low resolution OMI (Ozone Monitoring Instrument) data sets both use a TOMS-like (version 8) retrieval algorithm.
- The GOME, GOME2 and SCIAMACHY data sets all use the GODFIT retrieval algorithm (Lerot et al., 2010).
- As stated in Section 1.1 of the OMPS Algorithm Theoretical Basis Document (ATBD), the algorithm used for retrieving the OMPS TCO is adapted from the heritage TOMS version 7 retrieval.

Data set	Version	Period	Resolution (long x lat)
TOMS/OMI			
Adeos	7	1996-1997	1.25° × 1° ftp://toms.gsfc.nasa.gov/pub/adeos/
Source: https://data.ceda.ac.uk/badc/toms/data/adeos			
Earth Probe	8	1996-2005	1.25° × 1° ftp://toms.gsfc.nasa.gov/pub/eptoms/
Source: https://data.ceda.ac.uk/badc/toms/data/earthprobe/			
Meteor 3	8	1991-1994	1.25° × 1° ftp://toms.gsfc.nasa.gov/pub/meteor3/
Source: https://data.ceda.ac.uk/badc/toms/data/meteor3			
Nimbus 7	8	1978-1993	1.25° × 1° ftp://toms.gsfc.nasa.gov/pub/nimbus7/
Source: https://data.ceda.ac.uk/badc/toms/data/nimbus7			
Aura OMI, low resolution	8	2004-2012	1° × 1° http://tinyurl.com/p5qn9yd or http://tinyurl.com/qaeee7j
Source: https://disc.gsfc.nasa.gov/datasets/OMTO3e_003/summary			
Aura OMI, high resolution	8	from 2004	0.25 ° × 0.25° http://tinyurl.com/klwcfbn or http://tinyurl.com/mmccw8q
Source: https://disc.gsfc.nasa.gov/datasets/OMTO3e_003/summary			
SBUV/SBUV2			
Numbus Nimbus 7	8.6	1978-1990	overpass data ftp://toms.gsfc.nasa.gov/pub/sbuv/
Source: https://acdisc.gesdisc.eosdis.nasa.gov/data/Nimbus7_SBUV_Level2/SBUVN7O3.008/			
NOAA 9	8.6	1985-1998	overpass data ftp://toms.gsfc.nasa.gov/pub/sbuv/
Source: https://acd-ext.gsfc.nasa.gov/anonftp/toms/sbuv/			
NOAA 11	8.6	1988-2001	overpass data ftp://toms.gsfc.nasa.gov/pub/sbuv/
Source: https://acd-ext.gsfc.nasa.gov/anonftp/toms/sbuv/			
NOAA 16	8.6	2000-2003	overpass data ftp://toms.gsfc.nasa.gov/pub/sbuv/
Source: https://acd-ext.gsfc.nasa.gov/anonftp/toms/sbuv/			
NOAA 14	8.6	1995-2006	overpass data ftp://toms.gsfc.nasa.gov/pub/sbuv/
Source: https://acd-ext.gsfc.nasa.gov/anonftp/toms/sbuv/			
NOAA 17	8.6	2002-2013	overpass data ftp://toms.gsfc.nasa.gov/pub/sbuv/
Source: https://acd-ext.gsfc.nasa.gov/anonftp/toms/sbuv/			
NOAA 18	8.6	2005-2012	overpass data ftp://toms.gsfc.nasa.gov/pub/sbuv/
Source: https://acd-ext.gsfc.nasa.gov/anonftp/toms/sbuv/			
NOAA 19	8.6	2009-2013	overpass data ftp://toms.gsfc.nasa.gov/pub/sbuv/
Source: https://acd-ext.gsfc.nasa.gov/anonftp/toms/sbuv/			
ESA			
GOME	1.01	1996-2011	1° × 1° http://tinyurl.com/mw9wg6h
Source: https://earth.esa.int/eogateway/instruments/gome			
GOME2	1.00	2007-2012	1° × 1° http://tinyurl.com/q4ad575
Source: https://www.eumetsat.int/gome-2			
SCIAMACHY	1.00	2002-2012	1° × 1° http://tinyurl.com/lyz9alm
Source : https://www.sciamachy.org/products/index.php?species=O3			
OTHER			
NPP OMPS	1.0	from 2012	1° × 1° http://tinyurl.com/jvshwta
Source: https://disc.gsfc.nasa.gov/datasets/OMPS_NPP_NMTO3_L3_DAILY_2/summary?keywords=omps			

Table 1. The source data sets used to create version 3.4 of the NIWA-BS TCO database.

3 Determining corrections to TOMS and OMI data

First, the corrections required to the TOMS and OMI data sets are determined by comparing the satellite-based measurements with TCO measurements made by the global Dobson spectrophotometer and Brewer spectrometer networks. A map showing the locations of the Dobson and Brewer sites is available at https://woudc.org/data/stations/index.php?lang=en. We assumed in all cases that the Dobson and Brewer spectrophotometer data submitted to the WOUDC were the highest quality data available, that all possible corrections to improve the quality of the data had been made prior to submission, and that the measurements from these two networks were unbiased with respect to each other. Assumed uncertainties on the measurements from these two networks are presented below.

While TOMS and OMI data are provided in gridded data files, the original overpass data~~provide a higher quality data set~~, unlike the gridded data set where several measurements from different times through the day might be averaged within a grid cell, provide location-specific measurements that are more suitable for comparison with the ground-based measurement networks~~.~~, i.e. the overpass measurements are specific to a latitude, longitude and date/time. To this end, overpass data from the four TOMS instruments and from the OMI instrument were obtained from the GSFC (Goddard Space Flight Center) FTP server (ftp://toms.gsfc.nasa.gov). Dobson and Brewer measurements were obtained from the WOUDC (World Ozone and Ultraviolet Radiation Data Centre). Three-hourly means of the overpass data and direct-sun Dobson and Brewer TCO measurements were calculated for all sites for which Dobson and Brewer data were available. Exclusion of some of the Dobson and Brewer data, as discussed in Bodeker et al. (2001b), was required. Differences between 3-hourly means of ground-based and TOMS or OMI measurements were calculated. The uncertainties on the differences were calculated as the root sum of the squares of the uncertainties on the ground-based and satellite-based measurements (see Section 5).

The differences between the two data sets (satellite-based and ground-based) can be described as an offset and a drift i.e.

$$\Delta(t) = \alpha + \beta t \tag{1}$$

where t is the time in decimal years and α and β are fit coefficients, denoting the offset and drift respectively, to be determined through a regression model fit to the differences. Because the offset and drift between the two data sets is likely to depend on season and location, the α and β coefficients are expanded in a Fourier series to account for the seasonality (see e.g. Bodeker et al., 1998) and then further expanded in spherical harmonics to account for the latitudinal and longitudinal structure in the difference field. Based on theoretical expectations and past experience (Bodeker et al., 2001b) we assume that the differences do not depend on longitude. Under this assumption, the spherical harmonic expansions reduce to Legendre polynomials. The α coefficient then takes the form:

$$\alpha = \sum_{l=0}^{N_{L,\alpha}-1} L_l(\theta) \left(\alpha_{l0} + \sum_{f=1}^{N_{F,\alpha}} \alpha_{lf,\sin} \sin(2\pi f t) + \alpha_{lf,\cos} \cos(2\pi f t) \right) \tag{2}$$

where L_l denotes the l^{th} Legendre polynomial and θ is the co-latitude (90° - latitude). A similar expansion is made for β. The choice of $N_{L,\alpha}$, $N_{F,\alpha}$, $N_{L,\beta}$, and $N_{F,\beta}$ is somewhat arbitrary; the values need to be set sufficiently high to capture the seasonal and latitudinal structure in α and β but not so high as to over-fit the data and thereby introduce unrealistic structure

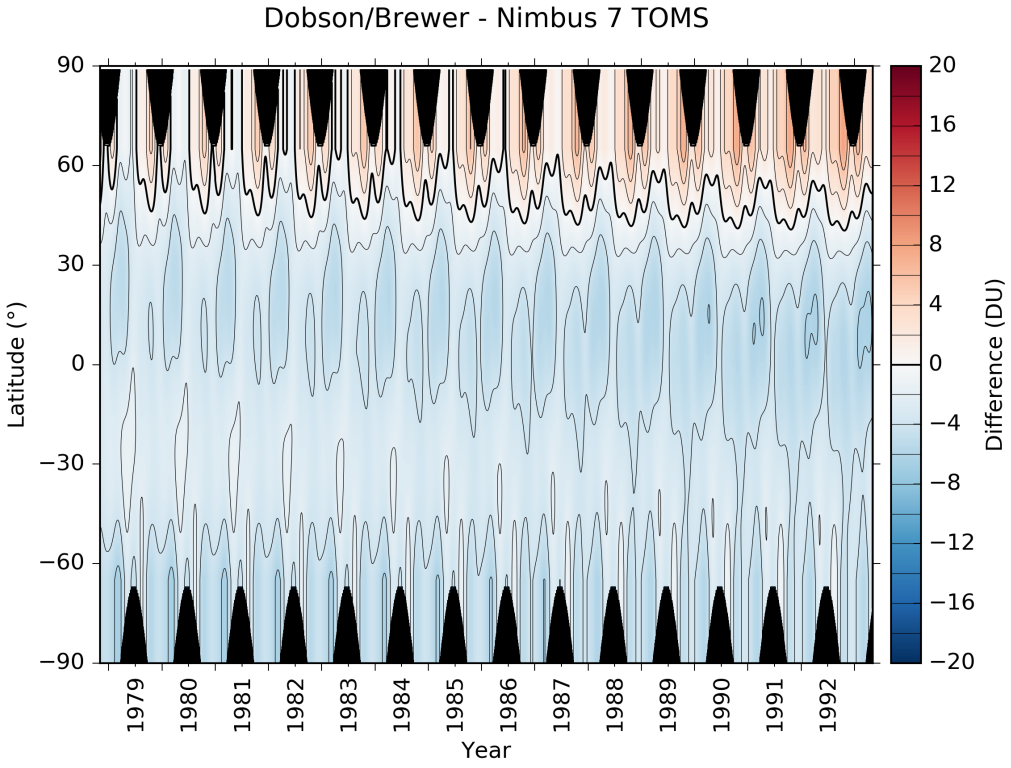

Figure 2. The results obtained by fitting Eq. 1 to the differences between Dobson/Brewer ground-based TCO measurements and Nimbus-7 TOMS overpass TCO measurements (ground-based minus satellite). Regions shaded in black denote the polar night where neither ground-based or space-based measurements are possible. The thick black line denotes the zero contour. Differences are shown in Dobson Units (DU; 1 DU = 2.69×10^{16} molecules/cm^2).

into the statistically modelled difference field. Visual inspection of a wide variety of different choices of $N_{L,\alpha}$, $N_{F,\alpha}$, $N_{L,\beta}$, and $N_{F,\beta}$ led to choice of (4,4,3,0) for the statistical model of the TOMS and OMI difference fields against the Dobson and Brewer networks where the Dobson and Brewer network is sparse, and (8,4,3,0) for differences from all other satellite-based data sources which are compared against the more dense, corrected, TOMS/OMI data set (see below). The choice of $N_{F,\beta} = 0$ is equivalent to there being no seasonal dependence in the drift. The resultant statistically modelled difference field is a compromise between high accuracy and low complexity or, equivalently, a compromise between simulating only meaningful structure in the difference field and avoiding over-fitting.

To avoid anomalous behaviour in the fit, which typically occurs at high latitudes and in regions where there are no satellite/ground-based difference pairs (e.g. during the polar night), the region between 80° and the pole is populated with difference values of zero for one month either side of the winter solstice. An example of one such fit is shown in Fig. 2. The morphology of this Dobson/Brewer - Nimbus-7 TOMS difference field is similar to that shown in Fig. 2 of Bodeker et al. (2001b) but with smaller

differences resulting from the use of Legendre expansions in latitude, which better accommodate hemispheric asymmetries, rather than a truncated polynomial expansion used in the earlier study. Similar $\Delta(t,\theta)$ difference fields (not shown) were statistically modelled for Adeos TOMS, Earth Probe TOMS, Meteor-3 TOMS, and Aura-OMI. Corrected TCO measurements for each of these data sets were calculated as follows:

$$TCO_{corr}(t,\theta,\phi) = TCO_{uncorr}(t,\theta,\phi) + \Delta(t,\theta) \quad (3)$$

where ϕ is the longitude.

4 Determining corrections to all other data sets

The TOMS and OMI grids, corrected for their offsets and drifts against the ground-based Dobson and Brewer measurements, now form the basis to correct the other data sets listed in Table 1. Differences between 1° zonal means from the combined corrected TOMS/OMI data and from the remaining data sets are calculated individually for each data set. The differences are then used as input to the regression model described in Eq. 1. There could be a danger here that in the case of biased longitudinal sampling by one satellite compared to another, that the zonal means would be biased but without these differences arising from any intrinsic biases between the satellite-based measurements. Only the SBUV measurements were sparse and corrections for this potential sampling bias were derived as discussed below.

If more than one TOMS/OMI meridional transect of zonal means is available for a given day, then all available difference values are passed to the regression model. As discussed in Bodeker et al. (2005), on 22 June 2003 a tape recorder failure on the ERS-2 satellite resulted in only a small portion of the Northern Hemisphere being sampled by the GOME instrument thereafter. To account for possible discontinuities in the difference field introduced by this anomaly, an additional basis function was included in the regression model for the TOMS/OMI-GOME differences, set to zero prior to 22 June 2003 and to 1 thereafter. The resultant fit to the differences between zonal means of TOMS/OMI and GOME is shown in Fig. 3.

The effects of the 22 June 2003 anomaly are clear in Fig. 3 with higher and more variable differences after 22 June 2003 than before. Statistically modelled difference fields, similar to that shown in Fig. 3, were generated for all non-TOMS/OMI data sets and extended, as required, to span the temporal coverage of each of those data sets to permit correction of the full data set. Because the combined TOMS/OMI record spans nearly the whole period (Nov 1978 to Dec 2016), extension into periods where TOMS/OMI data are not available is uncommon.

In addition to deriving the corrections for each data set listed in Table 1, the uncertainties on each of these corrections were also calculated since they contribute to the uncertainties of the respective data set as discussed in Section 5. The overall uncertainties on each of the source data sets are used to create an uncertainty weighted mean of all source data sets to produce the final TCO databases (see Section 6).

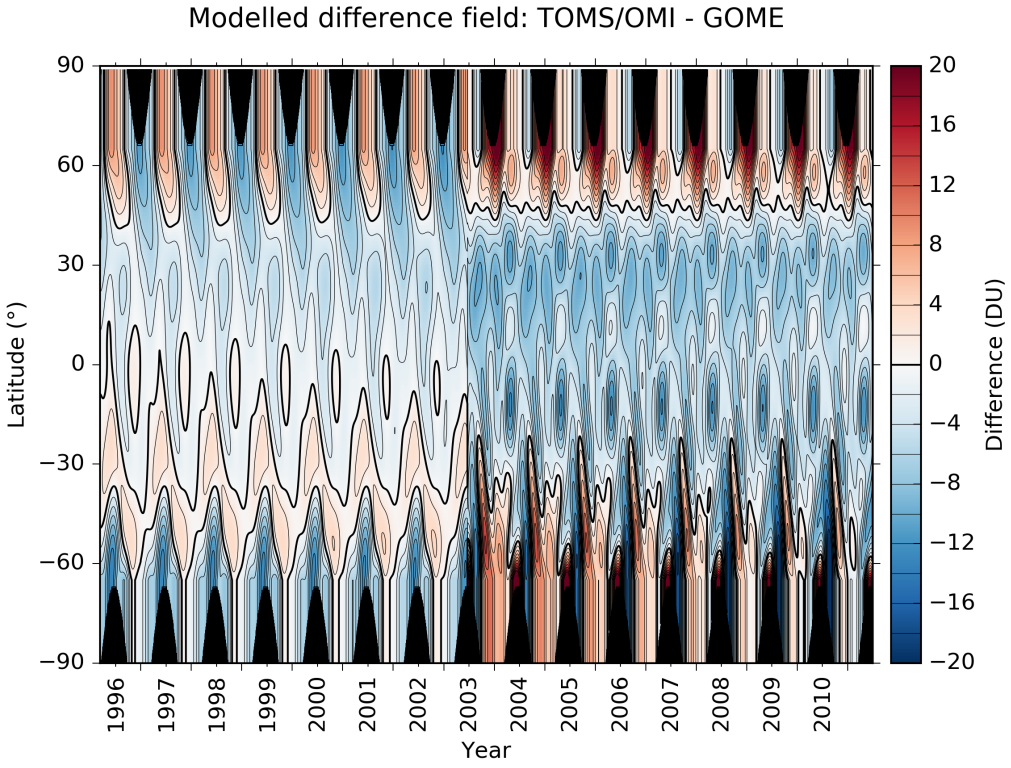

Figure 3. The results obtained by fitting Eq. 1 to the differences between corrected TOMS/OMI TCO zonal means and GOME TCO zonal means. An additional basis function is included to account for the 22 June 2003 anomaly. Regions shaded in ~~grey~~ black denote the polar night where space-based measurements in the UV/visible part of the spectrum are not possible. The thick black line denotes the zero contour.

5 Uncertainties on the source data sets

One attribute of this version of the NIWA-BS TCO database that differentiates it from previous versions is the provision of uncertainty estimates on each TCO value in the database. This development has been driven, in large part, by the requirements for a climate data record as stipulated in GCOS-143 (2010). Table 2 gives an overview of the literature on which we have based the uncertainty estimates of our source data sets. For the NPP-OMPS instrument, the uncertainty on each TCO measurement comprises both a static component (in DU), and a component that scales with the TCO i.e. is a percentage of the TCO. The relevant values (1.12 DU for the static component and 0.64 % for the component that scales with TCO) were derived from a linear fit to the data listed in Table 7.3-7 of ~~the Joint Polar Satellite System OMPS NADIR Total Column Ozone Algorithm Theoretical Basis Docume~~ Godin (2014).

The random uncertainties on the raw values listed in Table 2 are propagated through the analysis to result in an uncertainty estimate on the final product. When regression modelling the difference field between the ground-based Dobson/Brewer mea-

Data set	Random error TCO	Source
Dobson	1 %	~~Basher et al., 1980~~Basher and Bojkov (1995): Survey of WMO-sponsored Dobson Spectrophotometer intercomparisons
Brewer	1 %	~~Fioletov et al., 2005~~Fioletov et al. (2005): The Brewer reference triad
Adeos	2 %	~~Krueger et al., 1998~~Krueger et al. (1998): ADEOS TOMS Data Products User's Guide ~~http://tinyurl.com/k4mdlp~~
Earth Probe	2 %	~~McPeters et al., 1998~~McPeters et al. (1998): Earth Probe TOMS Data Products User's Guide ~~http://tinyurl.com/k~~
Meteor-3	3 %	~~Herman et al., 1996~~Herman et al. (1996): Meteor-3 TOMS Data Products User's Guide~~http://tinyurl.com/mm47~~
Nimbus-7	2 %	~~McPeters et al., 1996~~McPeters et al. (1996): Nimbus-7 TOMS Data Products User's ~~Guid macuv.gsfc.nasa.gov/~~
OMI	2 %	Bhartia (2002): OMI Algorithm Theoretical Basis Document Volume II, 2002~~http://tinyurl.com/knepwxk~~
all ESA	<1.7 % (SZA <80°) <2.6 % (SZA >80°)	GODFIT Algorithm Theoretical Basis Document (ATBD), 2013~~http://tinyurl.com/lg27pzd~~
NPP OMPS	1.12 DU, 0.64 %	Godin (2014): OMPS NADIR TCO Algorithm Theoretical Basis Document ~~, 2009 http://tinyurl.com/mprd4zp~~
all SBUV	5.0 DU	P.K. Bhartia: Personal communication, ~~2014.~~ 2014

Table 2. ~~Uncertainties~~ Typical uncertainties on the source data sets as reported in the references provided and as used in the construction of the TCO databases. Note that the uncertainty on any particular measurement may differ from the typical values quoted in this table.

surements and the TOMS/OMI overpass TCO measurements, the uncertainties passed to the regression model (Eq. 1) are:

$$\sigma_{\mathrm{diff}} = \sqrt{\sigma_{\mathrm{DB}}^2 + \sigma_{\mathrm{TOMS/OMIovp}}^2} \tag{4}$$

where σ_{DB} is the measurement uncertainty on the Dobson/Brewer measurements (1 %) and $\sigma_{\mathrm{TOMS/OMIovp}}$ is the measurement uncertainty on the TOMS/OMI overpass measurements.

The uncertainty on the modelled difference field is calculated using a Monte Carlo approach whereby the uncertainties on each difference pair are used to generate new estimates of the differences which then constitute a new data set of differences to which the statistical model is fitted. The process is repeated 100 times. The mean and standard deviation of the 100 resultant model fits provides the final difference field and its uncertainty ($\sigma_\delta \sigma_\Delta$). The uncertainties on the corrected TOMS/OMI values, calculated using Eq. 3, are then given by:

$$\sigma_{\mathrm{Corr}}(\theta,\phi,t) = \sqrt{\sigma_{\mathrm{uncorr}}(\theta,\phi,t)^2 + \sigma_\Delta(\theta,t)^2} \tag{5}$$

A similar procedure is used to propagate uncertainties in the corrections of the other satellite data sets against the corrected TOMS/OMI data sets. Recall that these corrections are based on comparisons of zonal means. To estimate the uncertainties on the zonal means, rather than taking the weighted mean of the single measurements, the unweighted arithmetic mean is calculated so that every measurement has the same weight. The zonal mean ZM, and its uncertainty σ_{ZM}, are then given by:

$$ZM = \frac{1}{N}\sum_i^N x_i \qquad \sigma_{ZM} = \frac{1}{N}\sqrt{\sum_i^N \sigma_i}\sqrt{\sum_i^N \sigma_i^2} \tag{6}$$

where N is the number of measurements in the zone, x_i are the measurements, and σ_i are the uncertainties on the measurements. The uncertainty on the zonal mean also needs to account for the effects of any undersampling. If the zonal profile of TCO is highly structured, perhaps as a result of planetary-scale waves, then, if a particular space-based instrument does not fully capture that structure, the uncertainty on the zonal mean will be higher than it would have been the case otherwise. This is primarily a concern for the sparse sampling by the SBUV instruments used to create the combined database. For the SBUV data sets, in addition to the zonal mean uncertainty calculated using Eq. 6, the potential uncertainty resulting from the sparse sampling was also accounted for. To estimate this additional uncertainty in the zonal means calculated using SBUV measurements an algorithm was developed to compare the zonal mean of a well-sampled zonal TCO profile with the zonal mean calculated using the same zonal profile but sampled at the SBUV measurement locations. The high spatial resolution (0.25°) OMI data set was used for this purpose. To estimate the potential zonal mean uncertainty for each day of the year resulting from the SBUV under-sampling, grids of high spatial resolution data from OMI were considered for the 10 years 2004 to 2013 and for 10 days before and after the day of interest totaling 21 days. This gives a sample data set of 210 data grids. For each SBUV data set available on that calendar day, and for each latitude, two zonal means are calculated, viz (1) the true zonal mean (ZM_{true}) calculated from the 1440 values comprising the zonal TCO profile at 0.25° resolution, and (2) the sub-sampled zonal mean (ZM_{sub}) calculated using only those OMI data at the locations of the SBUV measurements. For each latitude, 210 (from the 21 day window of 10 years of OMI data used) difference pairs of $ZM_{true} - ZM_{sub}$ can be calculated. If the SBUV sampling of the zonal mean was unbiased, all 210 values would be 0.0. The mean and standard deviation of these 210 values are then calculated and the standard deviation is used as an estimate of the SBUV sub-sampling uncertainty ($\sigma_{\mathrm{subsample}}$) which is specific to a particular SBUV instrument and depends on the year and latitude. An example of one such sub-sampling uncertainty field for the NOAA 19 SBUV instrument is shown in Fig. 4.

The sub-sampling uncertainty maximizes during periods when the zonal profile shows more complex structure i.e. typically in winter and spring when mid-latitude planetary wave activity maximizes. The sub-sampling uncertainty on the zonal means is added to the zonal mean uncertainty calculated using Eq. 6 as:

$$\sigma_{\mathrm{ZM_{SBUV}}} = \sqrt{\sigma^2_{\mathrm{ZM}} + \sigma^2_{\mathrm{subsample}}} \tag{7}$$

6 Creating the combined data set

To construct a single TCO field for each day, a weighted mean of all available corrected measurements in each grid cell is calculated. A grid of 1.25° longitude by 1.0° latitude was selected for the final product. The weights applied to the individual available TCO measurements are derived from the measurement uncertainties on each available measurement, viz:

$$\overline{TCO_{i,j}} = \frac{\sum_k w_{i,j,k} TCO_{i,j,k}}{\sum_k w_{i,j,k}} \qquad \sigma^2_{\overline{TCO_{i,j}}} = \frac{\sum_k w^2_{i,j,k} \sigma^2_{i,j,k}}{(\sum_k w_{i,j,k})^2} \tag{8}$$

where i and j are indices over latitude and longitude, k is an index over the measurements from different satellites in that cell, and the weights ($w_{i,j,k}$) are calculated as $1/\sigma^2$ where σ is the measurement uncertainty incorporating any additional uncertainty

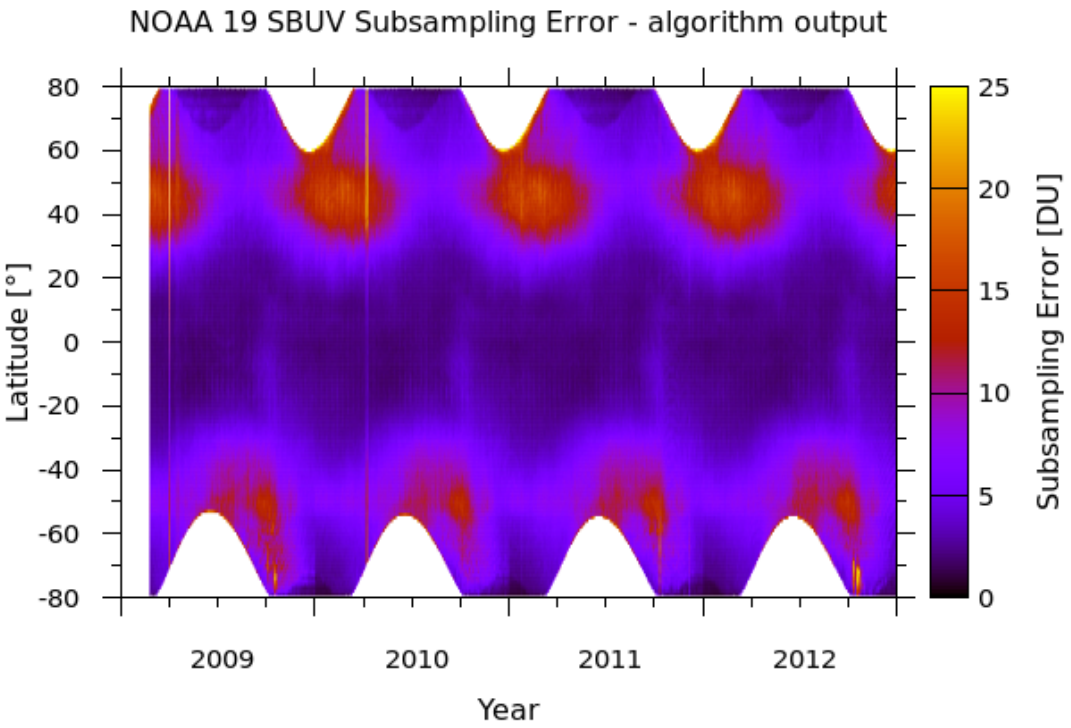

Figure 4. The additional uncertainty introduced to the SBUV zonal means as a result of under-sampling the zonal TCO profile. Regions shaded in white denote the polar night where space-based measurements in the UV/visible part of the spectrum are not possible.

introduced by corrections made to the original data. Unlike previous versions of this TCO database, the daily combined TCO fields are accompanied by fields of uncertainties and fields detailing the number of values that were averaged to produce the single combined value. An example of these three fields for one selected day is given in Fig. 5.

As expected, Fig. 5 shows that the uncertainty on TCO values decreases with an increasing number of source data sets. The regions of elevated uncertainty, sloping from north-west to south-east across the equator, arise from having only the OMI TCO values available to build the mean. Cyan regions in panel (c) show where additional Earth Probe TOMS data contribute (with a resultant reduction in the uncertainty) and regions in green where SCIAMACHY additionally contributes data, reducing the uncertainties in the resultant mean to less than 3 DU.

7 Validating the combined data set

This new NIWA-BS TCO database has been validated through comparisons with the WOUDC database and four additional independent TCO databases listed in Table 3. To account for different spatial resolutions, the NIWA-BS database was re-gridded to match the spatial resolution of each validation database. The differences reported in this section were calculated by

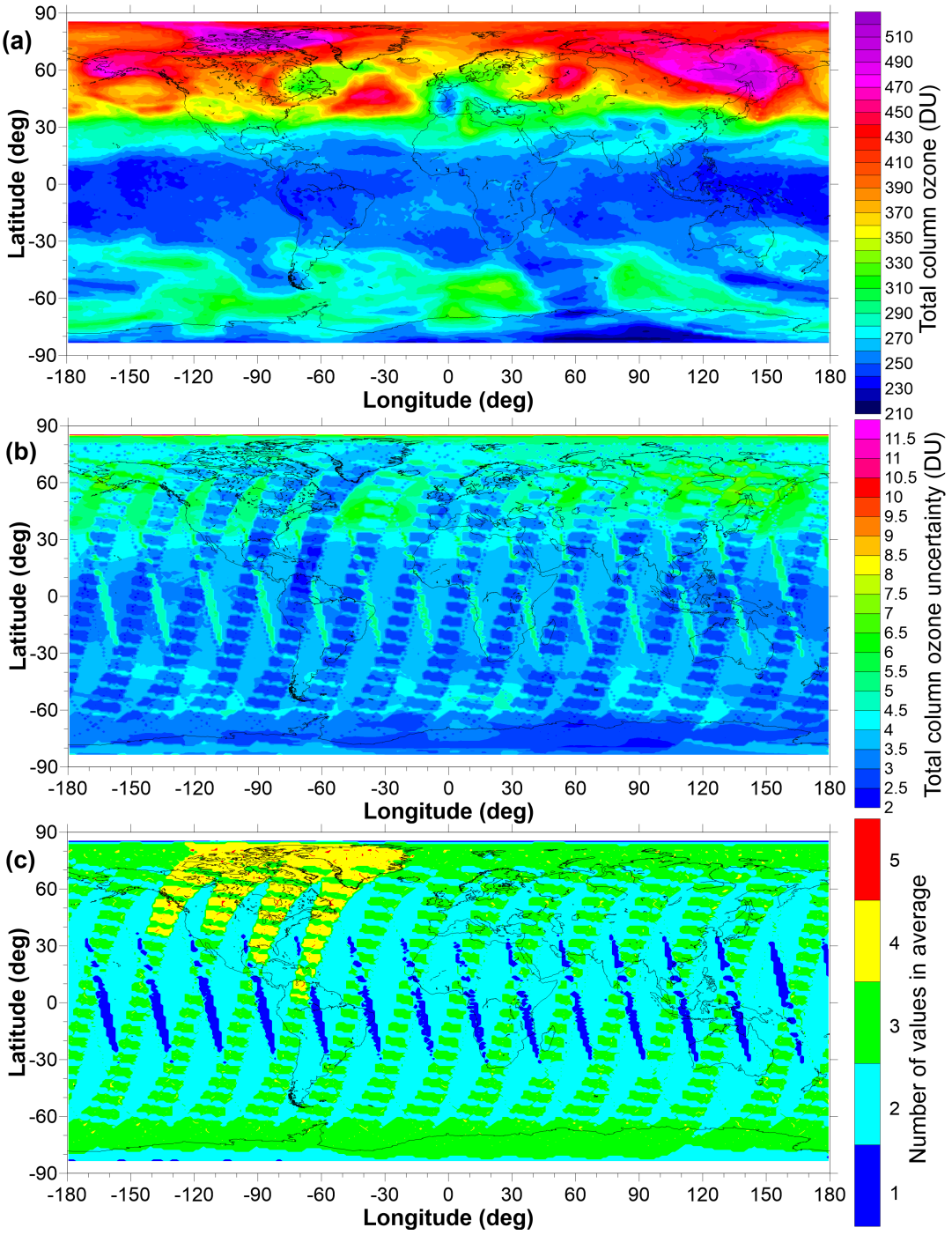

Figure 5. Example fields for 21 March 2005. (a) the TCO field, (b) the uncertainties on each value plotted in (a) and, (c) the number of values averaged to create the means plotted in panel (a). Regions shaded in white denote the polar night where space-based measurements in the UV/visible part of the spectrum are not possible.

Data set	Instruments	Record length	Reference
SBUV V8.6 NASA	BUV Nimbus-4, SBUV Nimbus-7, SBUV/2 NOAA 9 to 19	105/1978 to 71970 to 12/2013 2017	McPeters et al. (2013); Frith et al. (2014) http://tinyurl.com/m3m5qyn
Source: https://acd-ext.gsfc.nasa.gov/Data_services/merged/			
SBUV V8.6 NOAA	SBUV Nimbus-7, SBUV/2 NOAA 9 to 19	11/1978 to 12/2012 2015	Wild et al. (2012); McPeters et al. (2013) http://tinyurl.com/llbdssc
Source: https://acd-ext.gsfc.nasa.gov/Data_services/merged/instruments.html			
GSG Bremen	GOME, SCIAMACHY, GOME2	7/1995 to 712/2013 2016	Weber et al. (2013) http://tinyurl.com/l3gz6w6
Source: https://www.iup.uni-bremen.de/gome/wfdoas/			
ESA CCI	GOME, SCIAMACHY, GOME2	3/1996 to 6/2011	Lerot et al. (2014) http://tinyurl.com/q5h4bkj
Source: https://climate.esa.int/en/projects/ozone/			

Table 3. Sources and details of the independent data sets used to validate the NIWA-BS total column ozone database.

subtracting the validation values from the NIWA-BS values such that positive differences represent elevated ozone values in the NIWA-BS database compared to the validation databases. Fig. 6 shows the globally averaged area-weighted differences between the NIWA-BS database and the validation databases over the full time period.

The NIWA-BS database displays a small negative bias (-0.2±2.7 DU) against the global mean monthly means calculated from the Dobson and Brewer measurements obtained from the WOUDC. A slightly larger negative bias (-1.2±1.2 DU) is seen in comparison with the SBUV V8.6 NASA time series. The bias against the SBUV V8.6 data set produced by NOAA slightly more negative but not statistically significantly different from zero (-1.3±1.5 DU). The comparison against the GSG Bremen database suggests that the NIWA-BS time series exhibits a small anomalous downward trend starting around 2002 which is also reflected in the WOUDC comparison and in the ESA CCI comparison.

Seasonal mean differences between the NIWA-BS database and the five validation databases, as a function of latitude, are shown in Fig. 7. In general, the differences between the NIWA-BS TCO database and the validation databases are smaller than the uncertainties in the NIWA-BS database. This is not the case, however, in the high northern latitudes in winter where the NIWA-BS database shows statistically significantly smaller ozone values compared to the validation data sets. This results from larger differences in satellite measurements and ground-based measurements being inferred close to the region of permanent polar darkness where both satellite and ground-based measurements are scarce~~.~~, i.e. there were only 548 Dobson/Brewer - satellite difference pairs in the winter (DJF) Arctic (poleward of 60° N) on which to base the correction. A more in-depth comparison of the NIWA-BS database and the WOUDC database is presented in Fig. 8 where monthly mean zonal means (in

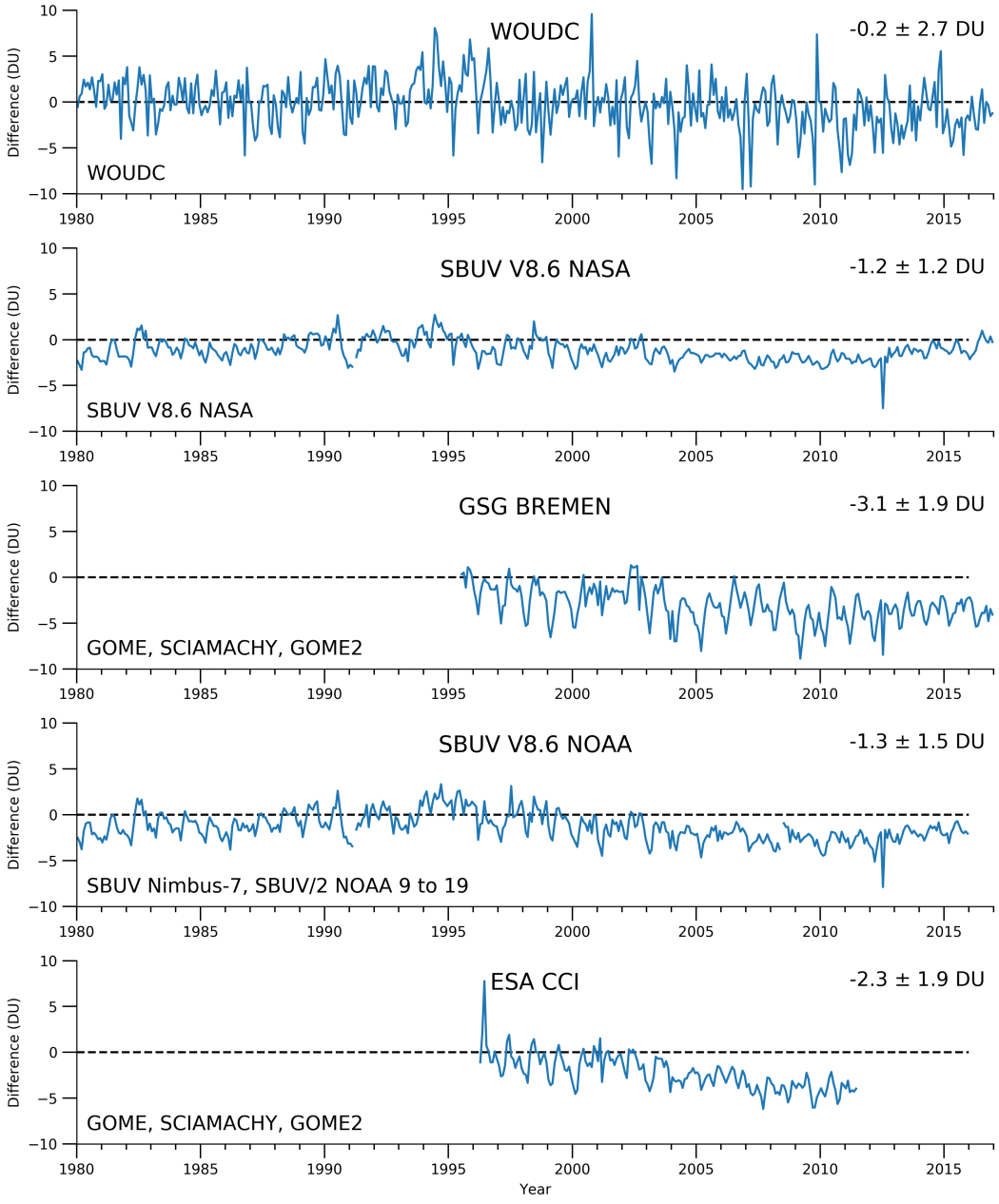

Figure 6. Area-weighted global mean monthly mean differences in TCO between the NIWA-BS database and the validation databases detailed in Table 3. The topmost panel shows the differences between the NIWA-BS database and the ground-based TCO database obtained from the WOUDC. The remaining four panels show differences against databases derived from space-based measurements. The statistics in the top right corner of each panel show the mean difference and standard deviation.

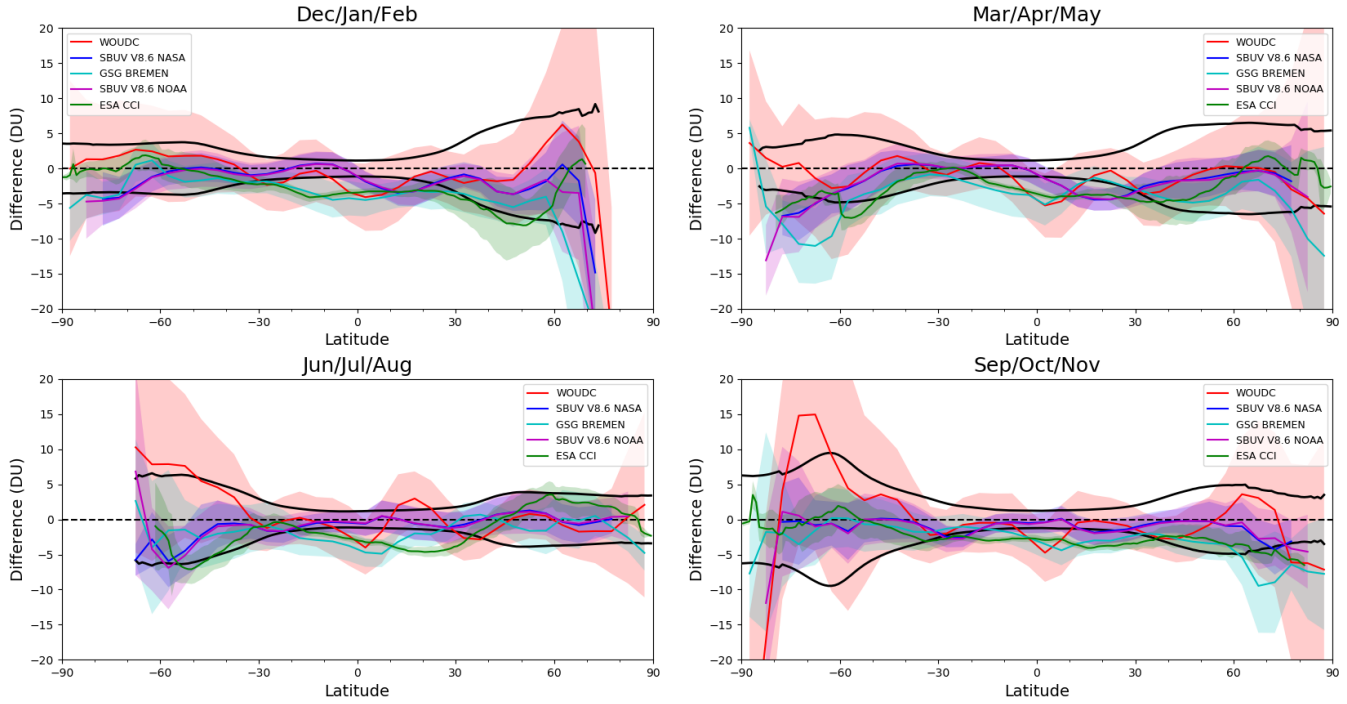

Figure 7. TCO differences (NIWA-BS TCO minus validation database) as a function of latitude plotted as seasonal means over the entire period of data available. The 1σ uncertainty range in the NIWA-BS TCO is shown in ~~grey~~black and the standard deviation on the differences between NIWA-BS TCO and each validation database is shown by the shading.

5° latitude zones) are differenced (NIWA-BS minus WOUDC). Over their full period of overlap, the mean difference between the data sets is -0.26 DU, with 95 % of the differences falling between -11.38 DU and 13.21 DU. Differences between the data sets are larger at higher latitudes. A consistent feature of the NIWA-BS TCO database across most years is an overestimation in TCO equatorward of the Antarctic and an underestimation of TCO close to the South Pole with respect to WOUDC.

Differences between monthly mean 5° zonal means from the NIWA-BS database and the NASA SBUV merged ozone database Version 8.6 are shown in Fig. 9. The mean difference in TCO across the full overlap period is -1.21 DU, with 95 % of the differences in the range -5.96 DU to 2.95 DU. The differences are smaller in magnitude than those shown in Fig. 8 and show smaller year-to-year variability, perhaps as a result of the more dense spatio-temporal sampling by the SBUV instruments compared to the ground-based instruments. Features common across most years are an underestimation of TCO in the northern sub-tropics during the first half of each year and underestimations just equatorward of the polar night in the Southern Hemisphere.

As the NOAA SBUV database is available at daily resolution like the NIWA-BS TCO database, daily differences between zonal means from these two databases are calculated and shown in Fig. 10. Over their full overlap period, the average difference

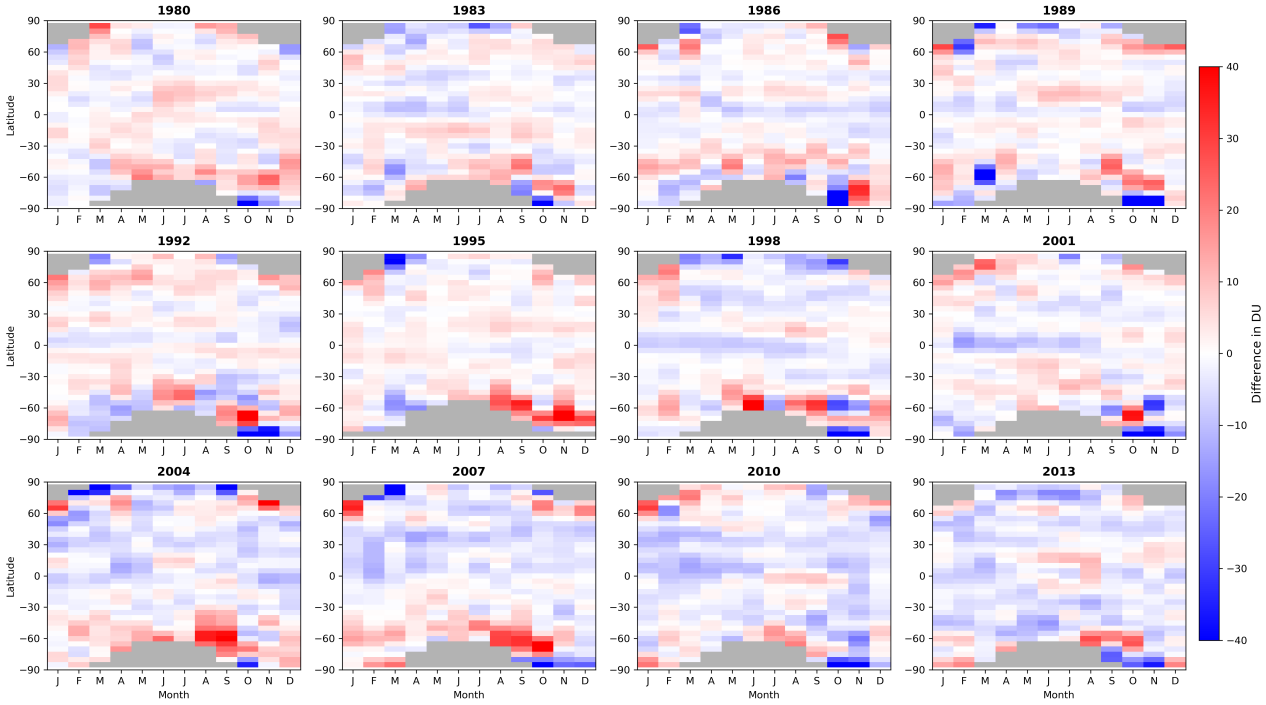

Figure 8. Monthly mean differences between NIWA-BS TCO and the WOUDC database for twelve selected years.

is -1.44 DU, with 95 % of the differences between -8.66 DU and 5.35 DU. Similar to the NASA SBUV comparisons, there appears to be a small underestimation in TCO over the northern sub-tropics in the first half of many years.

Differences between monthly mean 5° zonal mean TCO from the NIWA-BS database and the GSG Bremen TCO dataset which combined GOME, SCIAMACHY and GOME2 are shown in Fig. 11. There appears to be a consistent underestimation of TCO in the NIWA-BS database equatorward of the polar night from January to March in most years with respect to GSG Bremen. The mean difference between the databases is -3.14 DU, with 95 % of the differences lying between -9.79 DU and 2.81 DU.

Validation data from the ESA Climate Change Initiative (CCI) Level 3 TCO data set are available as monthly mean maps and differences between these monthly mean maps and NIWA-BS TCO are shown in Fig. 12. While there is significant spatial structure in some of the monthly difference fields, there is little structure that is consistent across multiple years. The mean difference is -2.36 DU, with 95 % of the differences lying between -10.63 DU and 6.34 DU.

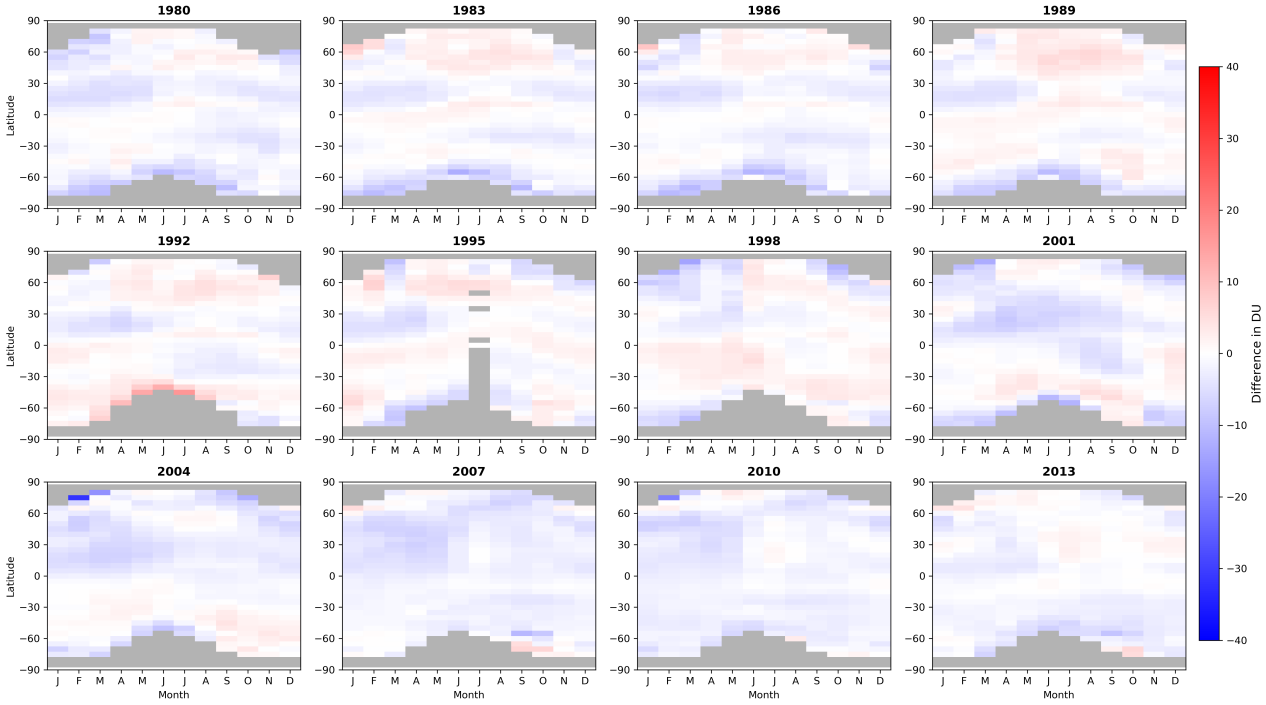

Figure 9. Differences between monthly mean zonal means calculated from the NIWA-BS TCO database and the NASA SBUV V8.6 database (NIWA-BS minus SBUV) for twelve selected years.

8 Calculation of monthly mean and annual mean fields

Monthly mean TCO fields at 1.25° longitude and 1° latitude resolution (the same resolution as the daily fields) have been calculated, together with their uncertainties. The algorithm was used to calculate the mean and its uncertainty from N measurements (x_i, i=1,...,N). The uncertainty on the mean is calculated in such a way that it depends on both the uncertainties on the measurements (σ_i) and on the variance in the measurements. First, a revised uncertainty for each datum is calculated to reflect the true confidence we have on each measurement as an estimator of the mean:

$$\sigma^2_{i-new\,i,new} = \sigma_i^2 + (x_i - x_{i,exp})^2 \tag{9}$$

where $x_{i,exp}$ is the 'expectation' value which is taken to be the unweighted mean of the available measurements. The mean is then calculated as:

$$\bar{x} = \frac{\sum_{i=1}^{N} w_{i,new} \times x_i}{\sum_{i=1}^{N} w_{i,new}} \tag{10}$$

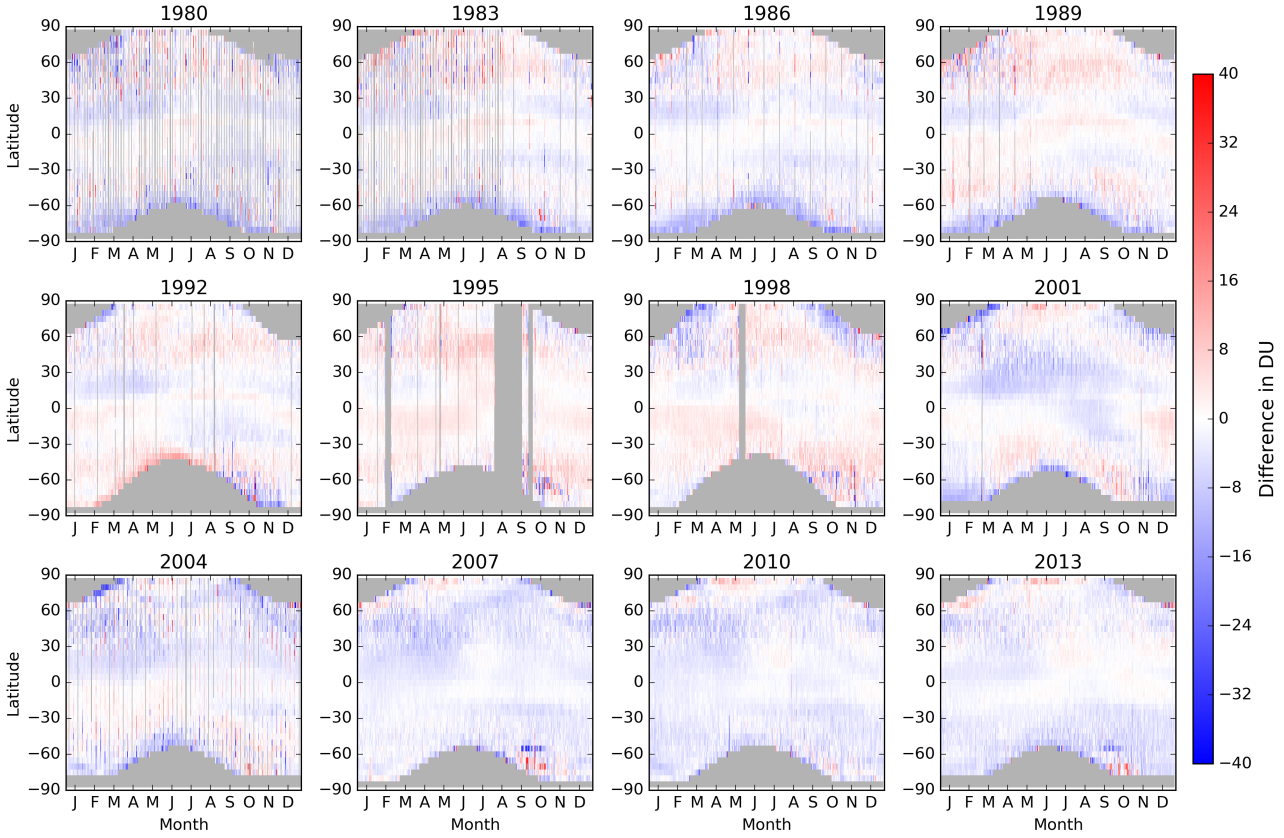

Figure 10. Daily differences between 5° zonal means calculated from the NIWA-BS TCO database and the NOAA SBUV V8.6 database for the same years as in Fig. 9.

where $w_{i,new} = 1/\sigma^2_{i,new}$ and the uncertainty is calculated as:

$$\sigma_{\bar{x}} = \sqrt{\frac{\sum_{i=1}^{N} \sigma^2_{i,new} \times w_i}{(NF-1) \times \sum_{i=1}^{N} w_i}} \tag{11}$$

where $w_i = 1/\sigma_i^2$ and NF is the degrees of freedom. In this case NF was taken to be $N-1$ to account, in part, for auto-correlation in the daily time series used to calculate the monthly and annual means. The monthly mean and annual mean TCO fields are provided as a component of the version 3.4 database.

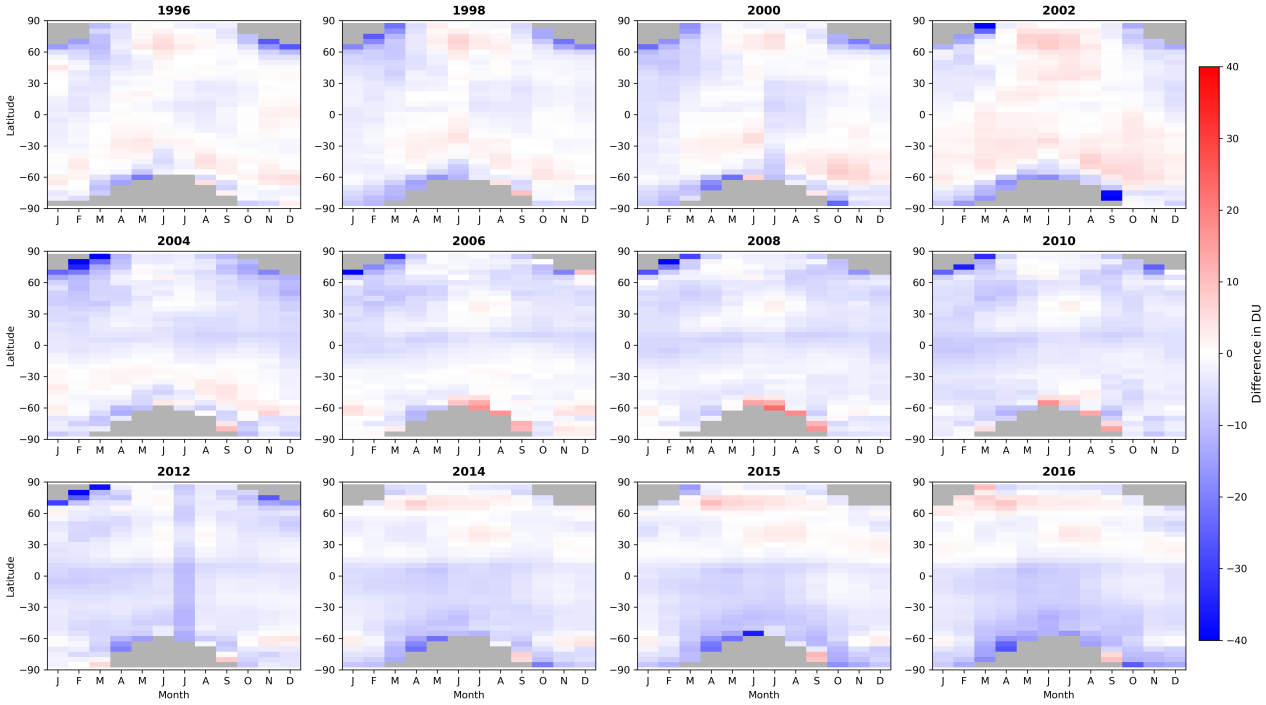

Figure 11. Differences between monthly mean 5° zonal mean TCO calculated from the NIWA-BS and GSG Bremen databases for twelve selected years (NIWA-BS minus GSG).

9 Creating the BS-filled Total Column Ozone database

For some applications, there is a need for gap-free TCO fields~~.~~, e.g., TCO fields for validating chemistry-climate models which generate TCO fields over the entire globe for each day of the year. To create a filled TCO database for a target day, the following steps are performed, each of which is detailed in sub-sections below.

1. A conservatively partially filled field is created (hereafter referred to as Field 1).
2. A machine-learning (ML) method is used to create a best estimate of the completely filled TCO field for the target day (hereafter referred to as Field 2).
3. The original unfilled TCO field, Field 1, and Field 2 are then 'blended' (using an algorithm described below) to generate the final filled field.

The result is a TCO field that replicates the original data where they are available and, where no data are available, transitions preferentially into the conservatively filled field (Field 1). Where the conservatively filled field has missing data, we transition into the ML-filled field (Field 2). Each 'transition' is achieved by way of the blending process described below.

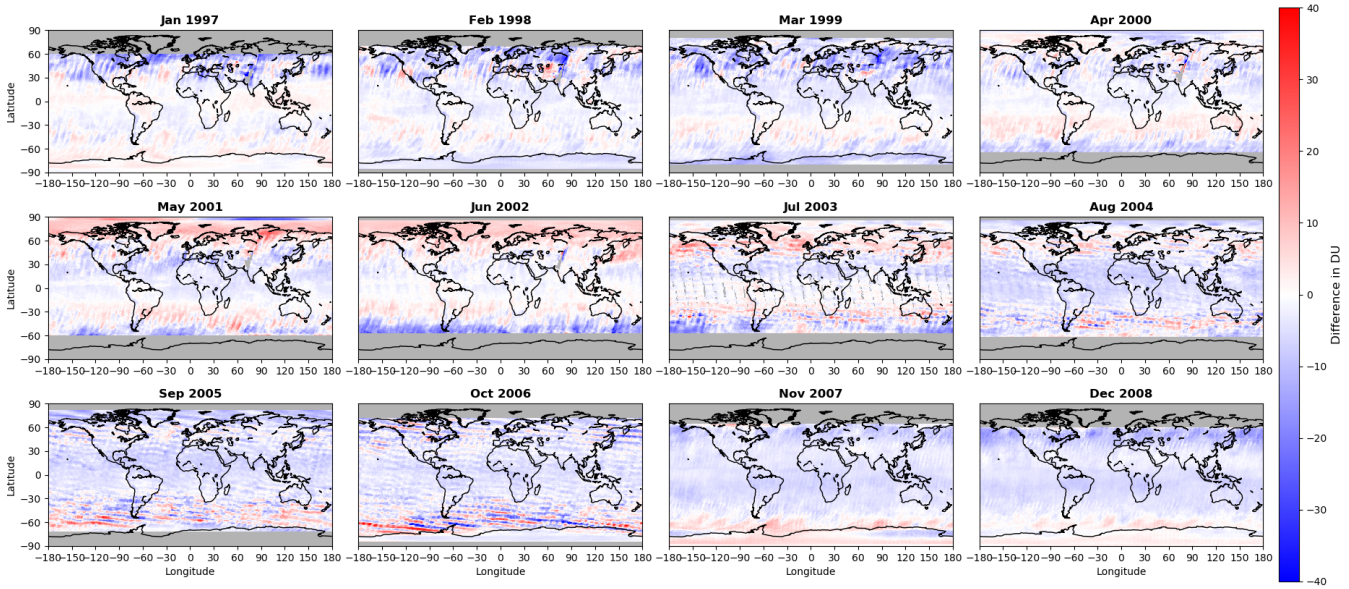

Figure 12. Differences between monthly mean TCO fields calculated from the NIWA-BS database and those available from the ESA CCI database, for twelve selected months/years.

9.1 The conservatively filled field - Field 1

First a spatial nearest neighbour interpolation is used to fill as many missing values as possible. This is done by searching for cells with null TCO values that are neighboured, either to the north and south, or to the east and west, by non-null values. If such a non-null pair is found, that pair of TCO values, together with their uncertainties, is used to estimate the interstitial value which is taken as the mean of the two neighbouring values. Preference is given to zonal nearest neighbors. The uncertainty on the interpolated values is calculated by adding in quadrature the uncertainties on the neighbouring values.

After doing the spatial nearest neighbour interpolation, cases are sought where, for a cell containing a null value on day N, there are non-null values in the same cell on day N-1 and day N+1. Temporal nearest neighbour interpolation between the previous and following day is done in the same way as described for the spatial nearest neighbour interpolation.

Following the spatial-temporal nearest neighbour interpolation, a more extensive longitudinal interpolation finds two non-null values at the same latitude that are separated by two or more null values with the constraint that the non-null values cannot be separated by more than 30° in longitude. Linear interpolation between the two non-null values, including an estimate of the uncertainties, is used to determine the interstitial values and their uncertainties. The spatial-temporal nearest neighbor interpolation, and the longitudinal interpolation, are repeated until no additional values are inserted into the grid. The result is a conservatively interpolated TCO field, still containing missing values, together with its original uncertainty field and traceable uncertainties on the newly interpolated values. A plot of the original TCO field, the conservatively filled field, and

the uncertainties on the conservatively filled field for day 3 of 1980 are shown in Fig. 13. The structure in the uncertainty field results from the propagation of uncertainties when calculating the conservatively filled field - larger uncertainties result when interpolated values are spatially far from available measurements.

9.2 The machine learning estimated field - Field 2

To create a completely filled TCO field for each day, a regression model, including an offset basis function, a tropopause height basis function, and a potential vorticity at 550 K basis function, is trained on a 'window' of data around the target date. The trained regression model is used to generate a filled TCO field on the target date.

The regression model is of the form:

$$TCO_{i,j} = \alpha(\theta,\phi) + \beta(\theta,\phi) \times TH_{i,j} + \gamma(\theta,\phi) \times PV550_{i,j} + R_{i,j} \tag{12}$$

where i,j subscripts denote indices over latitude (θ) and longitude (ϕ), TH and $PV550$ are 6-hourly tropopause height fields and 500 K potential vorticity fields, respectively, obtained from NCEP CFSR (National Centers for Environmental Prediction) CFSR (Climate Forecast System Reanalysis) reanalyses prior to 31 December 2010, and from NCEPCFSv2 reanalyses thereafter (Saha et al., 2010). α, β and γ are fit coefficients and $R_{i,j}$ are the residuals that remain due to variance that cannot be explained by the regression model.

As denoted in Eq. 12, the three fit coefficients depend on latitude and longitude. That dependence is captured by expanding the fit coefficients in spherical harmonics in a similar way as was done in Eq. 2, i.e.:

$$\alpha(\theta,\phi) = \sum_{l=0}^{N} \sum_{m=-l'}^{l'} \alpha_l^m Y_l^m(\theta,\phi) \tag{13}$$

where:

- θ is the co-latitude,
- ϕ is the longitude,
- N is a regression model parameter,
- $l' \leq l$ where the exact limit for l' is also a regression model parameter,
- α_l^m are the fit coefficients, and
- $Y_l^m(\theta,\phi)$ is the spherical harmonic function of degree l and order m.

The Y_l^m can be expressed as:

$$Y_l^m = \begin{cases} Y_l^0 & \text{if } m\text{=}0 \\ \sqrt{2} N_{(l,m)} P_l^m(\cos\theta)\cos(m\times\phi) & \text{if } m>0 \\ \sqrt{2} N_{(l,m)} P_l^{-m}(\cos\theta)\sin(m\times\phi) & \text{if } m<0 \end{cases} \tag{14}$$

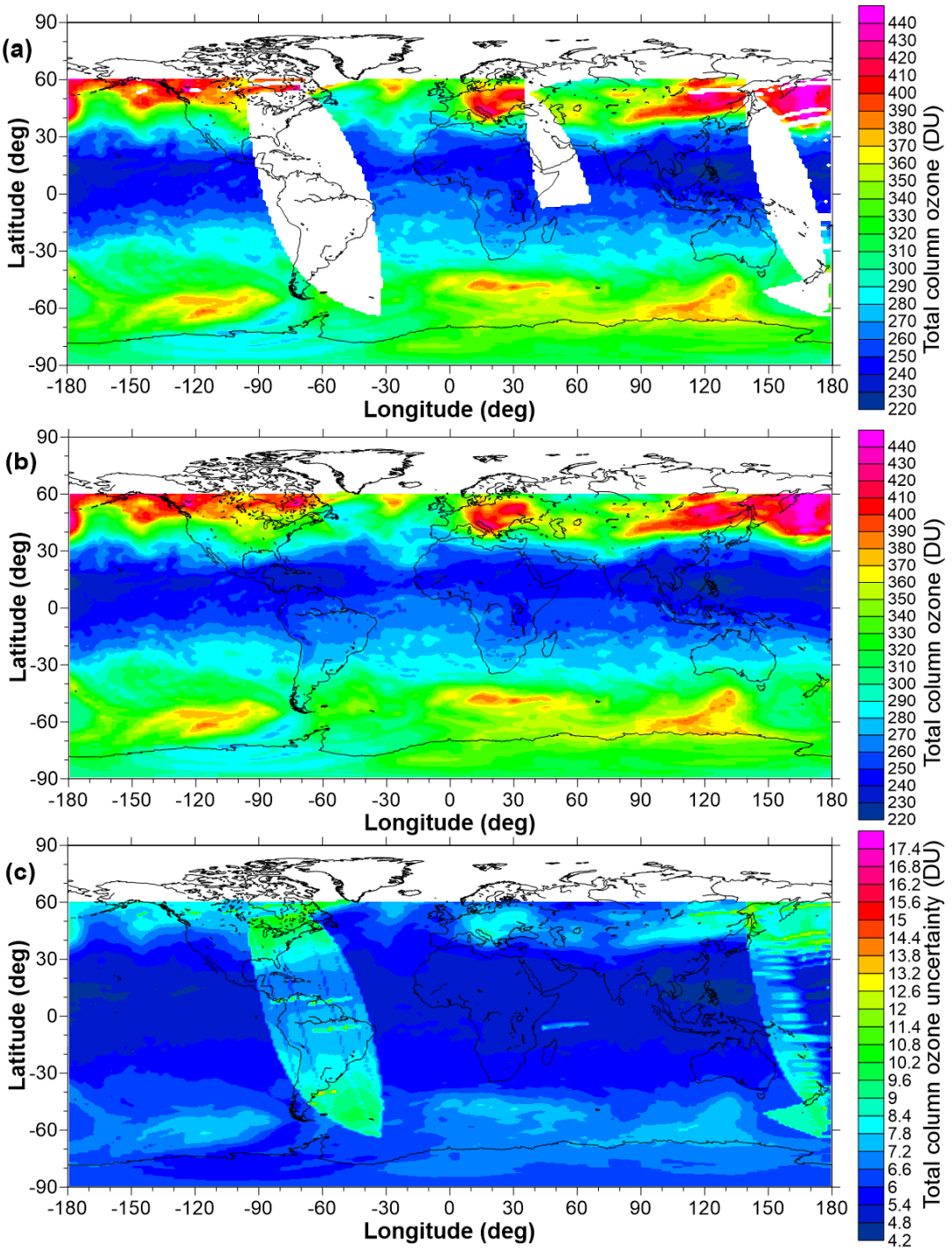

Figure 13. (a) The original unfilled TCO field on 3 January 1980, (b) the conservatively filled TCO field on the same day, and (c) the uncertainties on the conservatively filled field. White regions show where data are missing.

For the purposes of fitting Eq. 14 to the TCO fields, the normalisation constants ($\sqrt{2}N_{(l,m)}$) can be ignored as they are taken up into the fit coefficients. When fitting Eq. 12, the initial N value for α is set as 10 and for β and γ as 2. The maximum allowed N value for α is 10 and for β and γ is 5. The initial l' value for α was set at 5 (with a maximum allowed value of 5) and for β and γ as 2 (with maximum allowed values of 5). These initial values and limits on the spherical harmonics expansions were selected based on careful consideration of the scales of spatial structures in TCO fields, and on how the dependence of TCO on TH and $PV550K$ varies spatially. The training of the algorithm happens 'around' the target date as fields on neighbouring dates and years are used to establish the dependence of TCO on TH and $PV550K$. A search ellipse, initially extending 3 days either side of the target date, and 1 year either side of the target date, is defined to select TCO fields for the training where the ellipse is iteratively expanded until there are 20 TCO fields available for the training. The extension to neighbouring years is done because, in some cases, there are missing TCO fields in the current year such that a reliable fit of Eq. 12 cannot be performed. Within this search ellipse, the dependence of TCO on TH and $PV550K$ at a similar time of the year is expected to hold.

From these 20 TCO fields, to avoid excessive computational expense, only up to 20,000 data points are passed to the regression model by sampling every l^{th} value from all data available for training such that the total number of values passed is less than or equal to 20,000. The latitude and longitude of each ozone value is also passed to the regression model so that the associated TH and $PV550K$ values can be extracted. The times associated with the TCO fields are assumed to be local noon times (since most of the satellites making the underlying measurements were sun-synchronous satellites with an equator-crossing time close to solar noon). Therefore, the actual UTC time varies across the TCO field. The TH and $PV550K$ values are linearly interpolated to those exact UTC times.

Various 'versions' of the regression model are tested i.e. different basis functions are excluded/included; the offset (α) basis function is always included. In addition to switching different basis functions on/off, different values for N and l' are tested (perturbing these by ± 1 around their start value) but ensuring that the maximum allowed zonal and meridional expansions are not exceeded (see above). This results in many different possible constructs of the regression model. If any model, when evaluated over every latitude and longitude, results in a TCO value more than 10 $\%$ above the maximum measured TCO value passed to the regression model, or below 10 $\%$ less than the minimum value passed to the regression model, it is discarded to eliminate statistical models that significantly over-fit the TCO field. In addition to having a model with an excess of fit coefficients, over-fitting can also occur when anomalous values in the TH or $PV550$ fields result in excessively high or low TCO values being generated. This is why models that exclude/include these two basis functions are also tested. For all models that pass this initial test, a Bayesian Information Criterion (BIC; Liddle et al., 2007) score is calculated as:

$$BIC = M \times \ln(R^2/M) + NC \times \ln(M) \tag{15}$$

where M is the number of data passed to the regression model, R^2 is a modified sum of the squares of the residuals, and NC is the total number of coefficients in the fit. R^2 is modified to provide a strong disincentive for models generating values outside the range of measurements, i.e. where model values are below the minimum or above the maximum measurement passed to the

regression model, the residual is inflated exponentially to impose an additional cost on the model for generating values outside of the range of the data.

Typically, for each daily TCO field, several hundred fits of the regression model are performed to find the optimal model construct (minimum BIC). This optimal model is then used to generate the statistically modelled TCO field. The database of different regression models is also used to calculate the structural uncertainties that result from different possible choices of spherical harmonic expansions. The uncertainties that result from uncertainties on the regression model fit coefficients are also calculated. These two sources of uncertainty are then added in quadrature. The structural uncertainty statistics are calculated using only those regression modelled fields (out of the several hundred) that have the same sequence of basis functions switched on and off compared to the best fit. Examples of the unfilled TCO fields, the ML-modelled TCO fields and the uncertainty on the modelled TCO fields for days 304 and 305 of 1978 are shown in Fig. 14.

9.3 An algorithm for blending a primary and secondary TCO field

This section describes an algorithm to 'blend' some primary TCO field (hereafter Field A) with some secondary field (hereafter Field B) to create a single blended field (hereafter Field C) where the Field A values are preserved while smoothly transitioning to the Field B values. This algorithm is used below to combine the original TCO field and Field 1, and/or to combine Field 1 and Field2, and/or to combine the original TCO field and Field2; see Section 9.4.

If there is a null value in a cell in Field A and a non-null value in the same cell in Field B then a proxy value for Field A is found and combined with the value from Field B as follows:

$$C = W \times A_{proxy} + (1 - W) \times B_{value} \tag{16}$$

where C is the blended value, W is a weight calculated as detailed below, A_{proxy} is a proxy value for the missing value in Field A determined as detailed below, and B_{value} is the non-null value from Field B. To derive an A_{proxy} value, a box of size 41×41 cells is centred on the missing value and divided into six sectors each subtending an angle of $60°$ ~~sectors around that missing grid node are~~ . Each sector is scanned for non-null values ~~from Field Awhich are within 20 grid cells in the east-west and north-south directions. A~~ in Field A, and a weighted mean of ~~the 6 values from~~ those 6 ~~search sectors is then calculated , where the weighting is calculated as~~values is calculated these weights:

$$W_i = cos\left(\frac{D \times \pi}{2 \times 10^6}\right) \tag{17}$$

where D is the distance to the nearest non-null value in that sector measured in metres. The weight is set to zero when D is greater than 1000 km. A_{proxy} is then set to the weighted mean of the non-null values across all 6 sectors.

In calculating the blended value using Eq. 16, the weight (W) is calculated using the distance to the nearest non-null value across all 6 search sectors in Eq. 17. If no A_{proxy} value can be found, then W in Eq. 16 is set to zero. Standard error propagation rules are used to determine an uncertainty on the blended value. This process results in using values from Field A where they are available which then blend into values from Field B when they are not available.

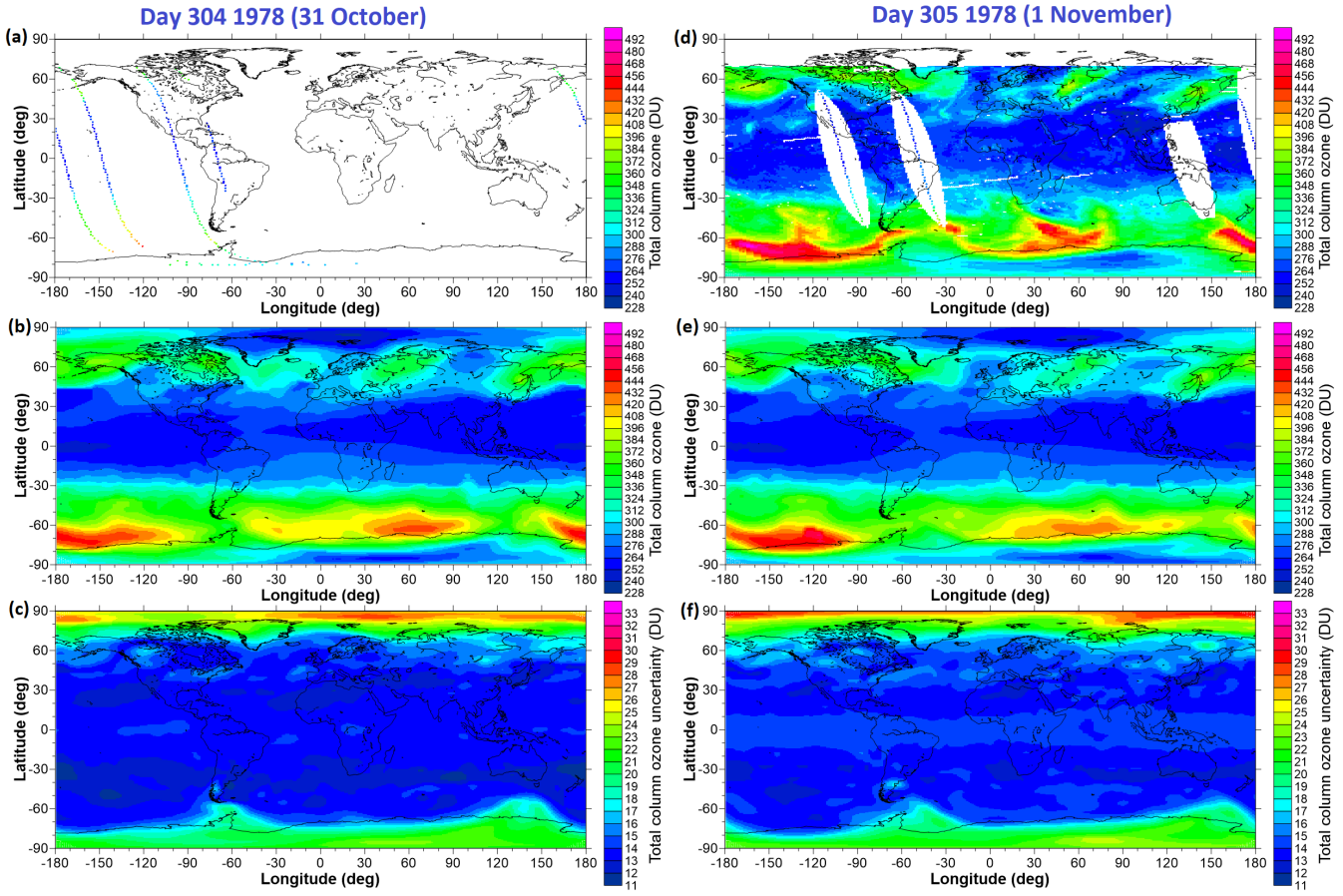

Figure 14. (a, d) The original unfilled TCO fields on 31 October 1978 and 1 November 1978, respectively; (b, e) the machine learning modelled fields; (c, f) the uncertainties associated with the machine learning modelled fields. On 31 October the optimal fit was obtained by expanding the offset basis function in spherical harmonics of degree 10 and order 5, expanding the tropopause height basis function in spherical harmonics of degree 4 and order 3, and the PV at 550 K basis function in spherical harmonics of degree 5 and order 3. For 1 November these expansions were (10,5) for the offset basis function (4,3) for the TH basis function and (5,4) for the $PV550K$ basis function.

9.4 Blending the original unfilled TCO field, Field 1, and Field 2 to construct the final filled field

For each day, the original TCO field, the conservatively filled field (Field 1) and the ML-modelled field (Field 2) are merged in such a way that the original values are preserved where they are available. Where they are not available, the filled values relax to the conservatively filled field where they are available. Where conservatively filled values are also not available, the filled values relax to the ML-modelled field.

The ML-modelled fields, because they are modelled on $PV550K$ and TH (which themselves can contain anomalous values), can occasionally display physically unrealistic spatial or temporal structures so, for any given day, to obtain a smoother ML-filled field, a (1,4,6,4,1) weighting of the five daily ML-filled fields, centered on the day of interest, is calculated.

For the final blended data product, several possibilities exist for any given day, i.e.:

- *None of the three fields are available:* In this case no final filled field is generated.
- *Only the ML-modelled field is available:* In this case the ML-modelled field (possibly spatially smoothed), and its associated uncertainties, become the final filled field for the target day.
- *Only the conservatively filled field and the ML-modelled fields are available:* In this case the conservatively filled field (Field 1) and the ML-modelled field (Field 2) are blended to create the final filled field.
- *All three fields are available:* In this case the conservatively filled field and the ML-modelled field are blended to create an intermediate field. The original TCO field and the intermediate field are then blended to create the final filled TCO field and its uncertainty.

An example of the original field for TCO on 31 October 1978, the conservatively filled field, final field obtained from the blending process, and the uncertainty on the final filled field are shown in Fig. 15 (the ML-modelled field for that day can be seen in Fig. 14).

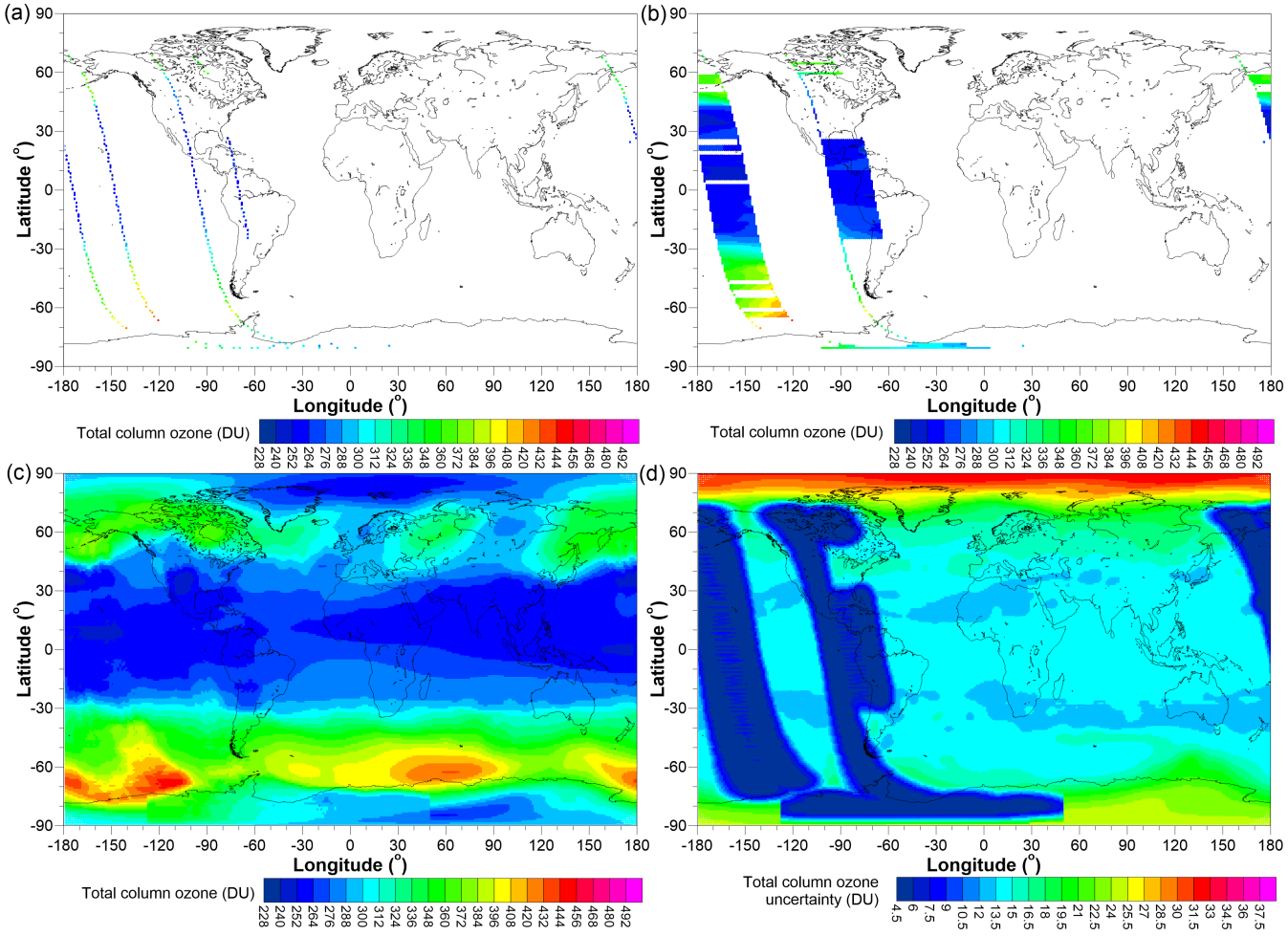

Figure 15. (a) The original unfilled TCO field on 31 October 1978, (b) the conservatively filled field, (c) the final filled field, and (d) the uncertainties on the final filled field.

10 Trend analysis

To prove the utility of the NIWA-BS TCO database, trends over the whole period have been diagnosed using the following regression model, described in detail in Bodeker et al. (2013):

$$
\begin{aligned}
\text{Ozone}(m,\theta,\phi) = & A(m,\theta,\phi) + && \text{Max Fourier} = 4 \\
& B(m,\theta,\phi) \times m/12 + && \text{Max Fourier} = 3 \\
& C(m,\theta,\phi) \times m_{\text{m=0 if year<2000}}/12 + && \text{Max Fourier} = 3 \\
& D(m,\theta,\phi) \times \text{QBO}(m) + && \text{Max Fourier} = 2 \\
& E(m,\theta,\phi) \times \text{QBO}_{orthog}(m) + && \text{Max Fourier} = 2 \\
& F(m,\theta,\phi) \times \text{ENSO}(m) + && \text{Max Fourier} = 1 \\
& G(m,\theta,\phi) \times \text{Solar}(m) + && \text{Max Fourier} = 0 \\
& H(m,\theta,\phi) \times \text{Pinatubo}(m+\Delta m) + && \text{Max Fourier} = 1 \\
& I(m,\theta,\phi) \times \text{El Chichón}(m+\Delta m) + && \text{Max Fourier} = 1 \\
& R(m,\theta,\phi)) && (18)
\end{aligned}
$$

where Ozone(m,θ,ϕ) is the regression modelled TCO in month m (m = 1 to $NY \times 12$ where NY is the total number of years of data) and at latitude θ and longitude ϕ. The monthly mean TCO values were calculated as detailed in Eq. 10 and Eq. 11. Equation 18 is fitted independently to the monthly mean time series at each latitude and longitude. A to I are the regression model coefficients calculated using a standard least squares regression (Press et al., 1989).

The first term in the regression model (A coefficient) represents a constant offset and, when expanded in a Fourier series, represents the mean annual cycle. In addition to the offset coefficient, each fit coefficient can depend on season, e.g., TCO trends vary with season. Therefore each coefficient is expanded in Fourier pairs as explained in Section 2.2 of Bodeker et al. (2015). The actual number of Fourier pairs for each regression coefficient is determined by finding the optimal set of expansions across all fit coefficients that minimizes a BIC as described above. The maximum number of Fourier pairs permitted for each regression model coefficient is listed in Eq. 18. The B coefficients diagnose the trend over the full period while the C coefficients diagnose the change in trend from 2000 onward. 1999/2000 was prescribed as the trend transition year as this is approximately when stratospheric chlorine and bromine loading peaked (Newman et al., 2007). We also wanted to ensure that the first trend period included data from the late 1990s as there was a greater likelihood of missing data from 1994 to 1998 and we wanted to avoid end-effect-biasing in the calculation of the trends. That said, the conclusions drawn below regarding changes in trend were found to be largely insensitive to the selection of the transition year within 2 years of the selected transition year. The QBO basis function was specified as the monthly mean 50 hPa Singapore zonal wind. The phase of the QBO varies with latitude and, to permit fitting of the phase, a second QBO basis function, mathematically orthogonalized to the first, was included in the regression model as was done in Bodeker et al. (2013).

The El Niño Southern Oscillation (ENSO), solar cycle, Mt. Pinatubo and El Chichón basis functions were the same as those used in ~~Bodeker et al. (2001a)~~Bodeker et al. (2001b). Note that the effects of the Pinatubo and El Chichón volcanic eruptions on TCO was delayed and so temporal offsets in those basis functions, denoted by the Δm, were permitted for up to two years where the Δm values for each basis function were optimised as part of the same BIC optimisation used to determine the optimal Fourier expansions for each regression model fit coefficient. A one term auto-correlation model was used to account for the effects of auto-correlation in the residuals ($R(m, \theta, \phi)$) when calculating the uncertainties on the fit coefficients.

The B coefficients, evaluated at day 90, 180, 270 and 360 through the year (noting that because these are Fourier expansions they can be evaluated at non-integral month values) are shown in Fig. 16. The spatial and seasonal pattern of ozone trends prior to 2000 is similar to numerous previous studies that have shown similar trend results. The C coefficients, evaluated on the same days as the B coefficients, are shown in Fig. 17. Note that Fig. 17 shows the **change** in trend compared to the trend shown in Fig. 16 i.e. areas midway between blue (more negative trends than before 2000) and red (more positive trends than before 2000) indicate no change in trend before and after 2000. Actual trends since 2000 obtained by adding the trends shown in Fig. 17 to the trends shown in Fig. 16 are shown in Fig. 18. Trends are generally seldom statistically significantly different from zero at the 2σ level. Regions of the southern middle and low latitudes continue to show declining TCO while selected regions over Antarctica show positive TCO trends.

While, since 2000, TCO trends have largely shown a positive change compared to pre-2000 trends (particularly over Antarctica as the ozone hole recovers due to effects of declining stratospheric concentrations of ozone depleting substances), around the middle of the year a large negative change in trend in excess of -1 DU/year is seen in the Southern Hemisphere sub-tropics. The largest of these negative anomalies occurs at 19.5°S, 164.375°W and therefore Fig. ~~18~~ 19 shows the July mean TCO, regression model fit for July, and the contribution of the offset and trend basis function evaluated at this location. The strong negative change in trend suggested by the regression model is clearly apparent in the observations. This decline in Southern Hemisphere sub-tropical TCO seen here is consistent with other reports of continued declines in tropical ozone after 2000 (Ball et al., 2018). Further research is required into the mechanisms that are driving such decreases in tropical TCO in spite of the effectiveness of the Montreal Protocol and its Amendments and Adjustments in reducing the stratospheric burden of halogens.

11 Conclusions

This paper presents the construction of a new version (V3.4) of the NIWA-BS TCO database and the developement of the gap-free BS-filled TCO database. To the extent possible, we have followed the recommendations of GCOS-143 (2010) in establishing a fundamental climate data record for TCO, in particular paying specific attention to tracing all sources of uncertainty through to the final data product. Making the uncertainties available per datum presents a major advancement from the last version of the NIWA-BS TCO database. Comparisons of the NIWA-BS TCO database against the WOUDC database and four independent multi-satellite databases show generally small differences that are within the uncertainties of the NIWA-BS TCO database. The BS-filled TCO database provides gap-free TCO fields for each day that have been created from a machine-learned algorithm that uses tropopause height and/or potential vorticity at 550 K fields as estimators of the spatial and temporal

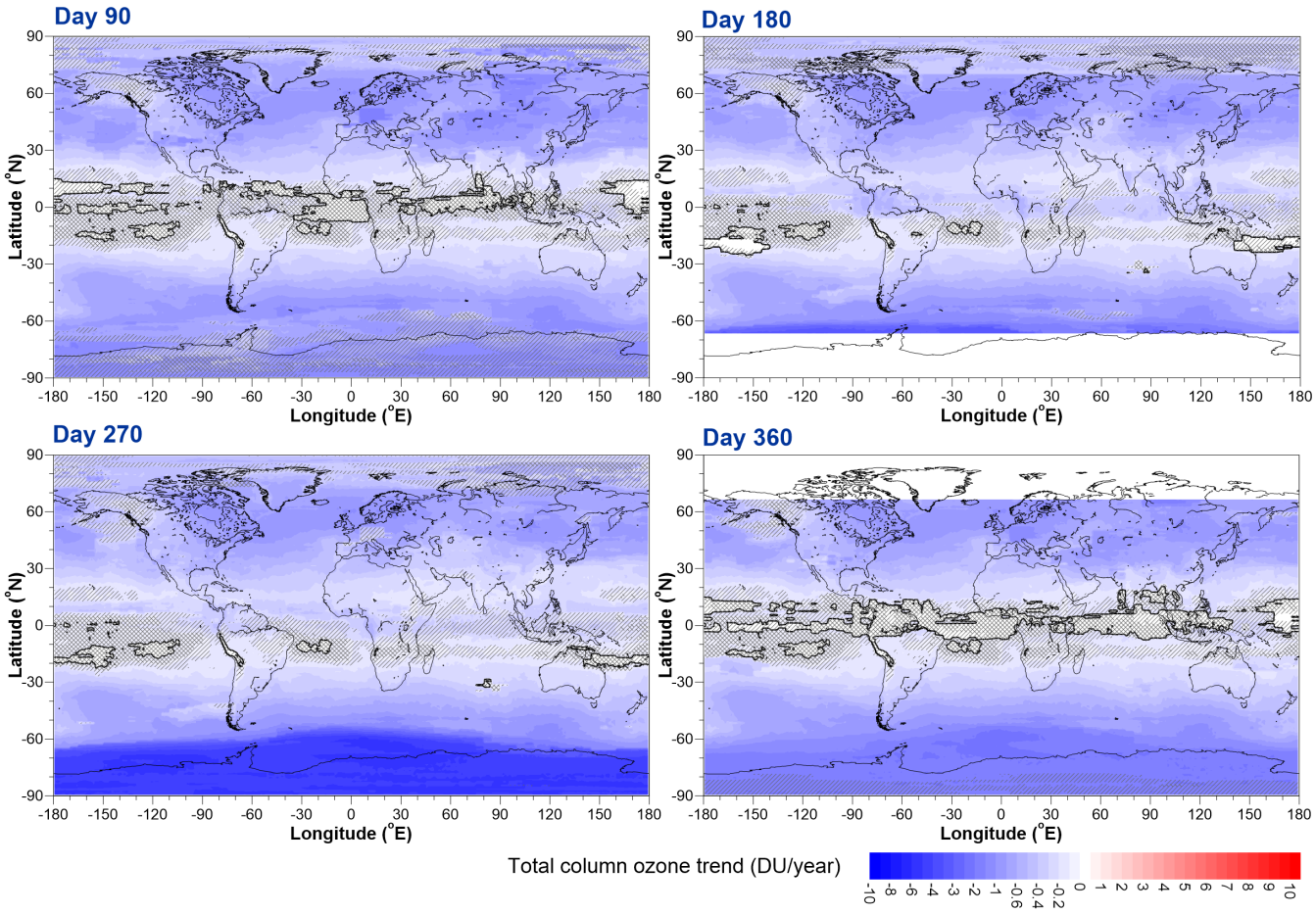

Figure 16. TCO trends for the period 1979 to 2000 evaluated at (a) day 90, (b) day 180, (c) day 270, and (d) day 360. The white areas at high latitudes in (b) and (d) result from there being insufficient data to establish meaningful trends during the polar night periods. The solid black line shows zero trend. Double hatched regions show where trends are not statistically different from zero at the 1σ confidence level, while single hatched regions show where the trend is not statistically different from zero at the 2σ confidence level.

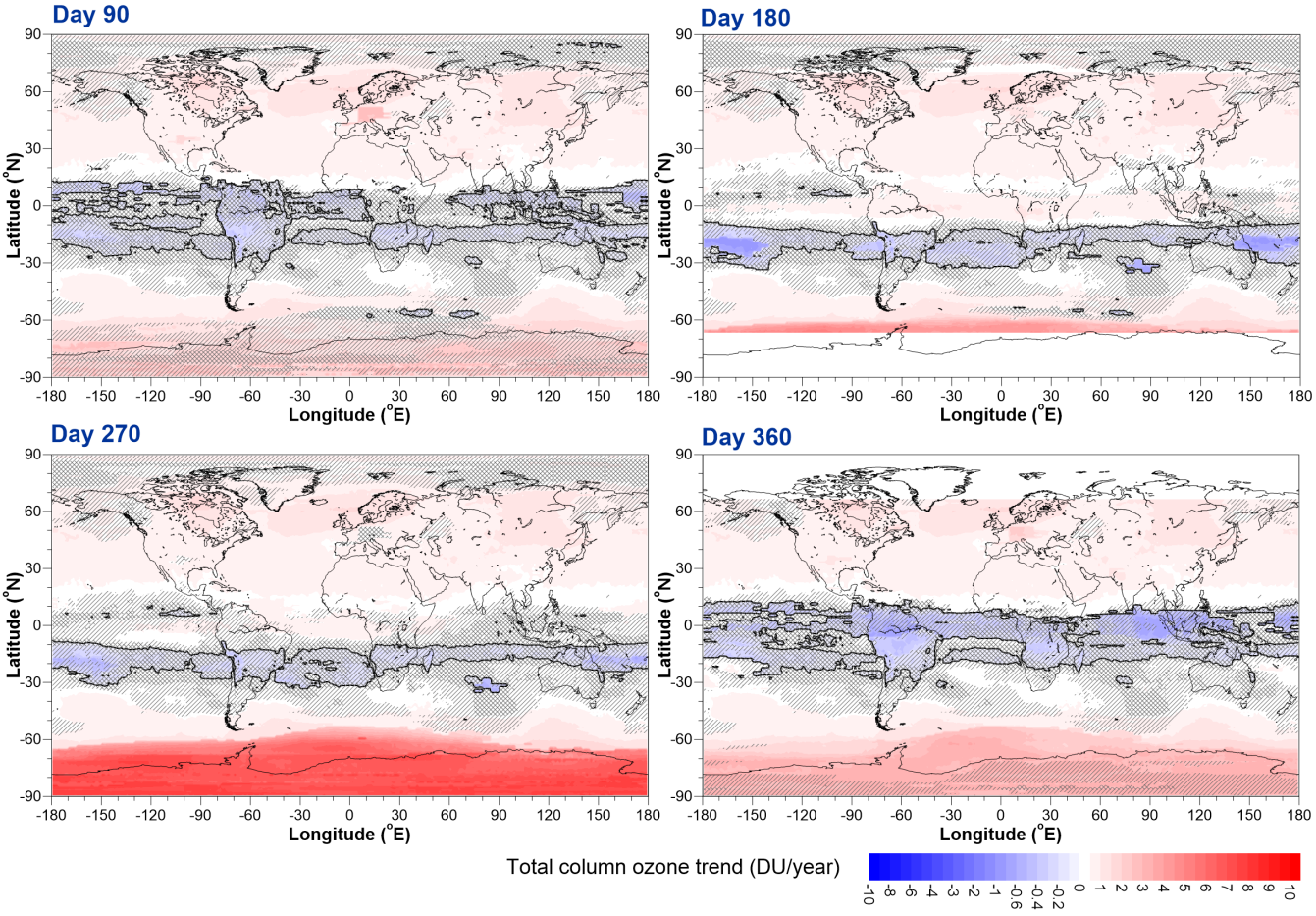

Figure 17. The change in TCO trends after 2000 evaluated at (a) day 90, (b) day 180, (c) day 270, and (d) day 360. The solid black line shows zero trend. The same hatching regimen as used in Fig. 16 is used here.

variability in the TCO fields. Finally, an analysis of trends in unfilled monthly mean TCO fields showed that while many regions of the globe that had been experiencing negative trends in TCO prior to 2000 are now seeing positive TCO trends, there are regions in the tropics (with a bias towards the Southern Hemisphere) where trends since 2000 have become more negative and over limited regions of the tropics, this decline is statistically significant. The cause of this ongoing negative trend has not been diagnosed here and requires further investigation.

12 Data availability

The NIWA-BS TCO database (doi:10.5281/zenodo.1346424, Bodeker et al. (2018)) and the BS-Filled TCO database (doi:10.5281/zenodo.3908787, Bodeker et al. (2020)) are available from http://www.bodekerscientific.com/data/total-column-

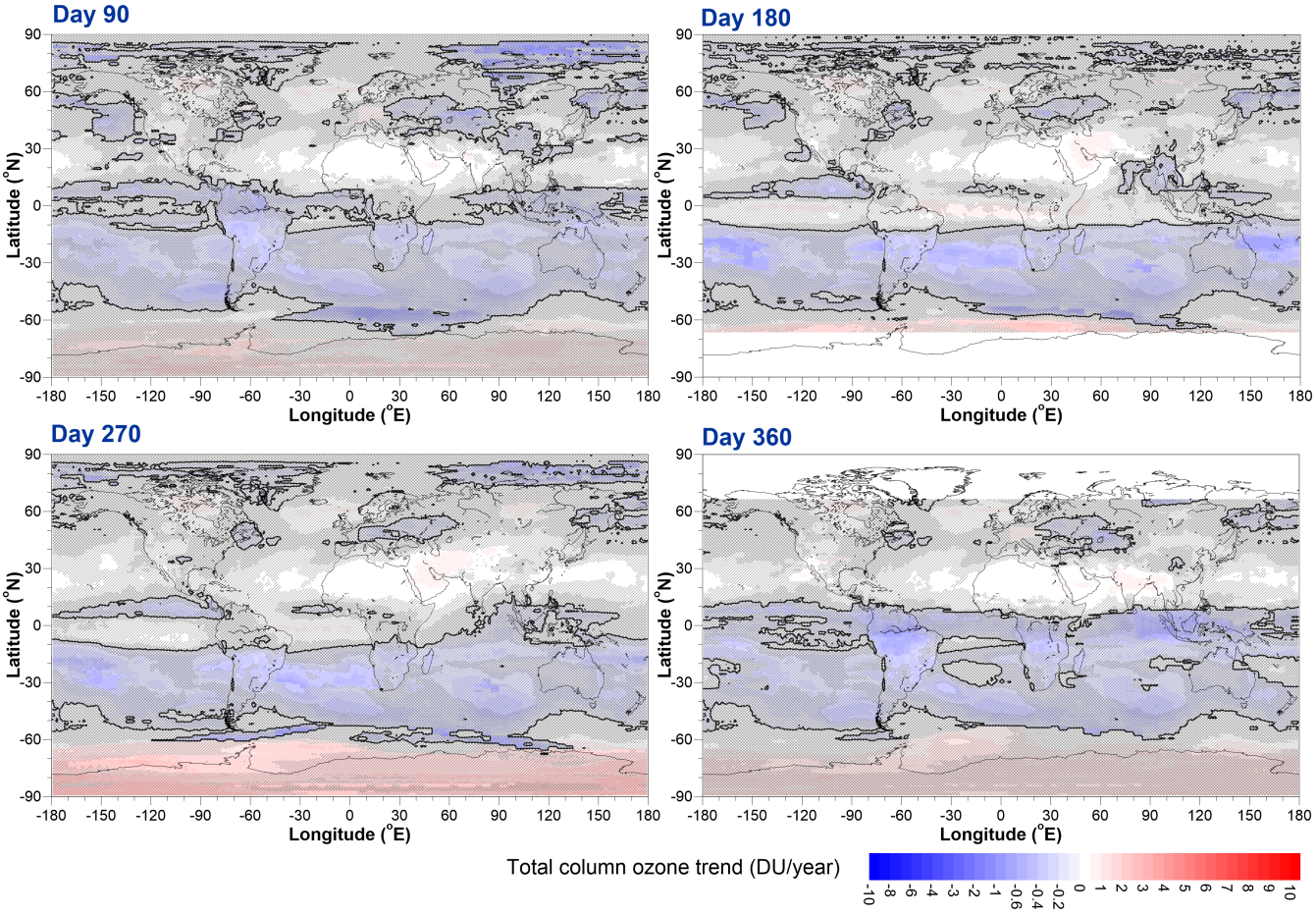

Figure 18. Inferred trends in TCO 2000 evaluated at (a) day 90, (b) day 180, (c) day 270, and (d) day 360, obtained by adding the B and C regression model fit coefficients. The solid black line shows zero trend. The same hatching regimen as used in Fig. 16 is used here.

ozone and from the zenodo archive. Both databases are freely available for non-commercial purposes and are provided as netCDF files.

Author contributions. GEB wrote much of the code for processing the TCO data files into a common format, the code for filling missing data, performed the trend analysis, and wrote the majority of the paper. JN obtained all of the uncertainty information for the different data sets and wrote the code for tracing those uncertainties through the processing chain. JST assisted with the data analysis, the generation of figures, and the writing of the paper. SK ran and debugged the code for applying the corrections to the 17 different TCO data sets and assisted with the writing of the paper. AS wrote some of the code for homogenising the non-TOMS/OMI data sets. JL wrote much of the C++ code for deriving the statistical correction fields between the non-TOMS/OMI data sets.

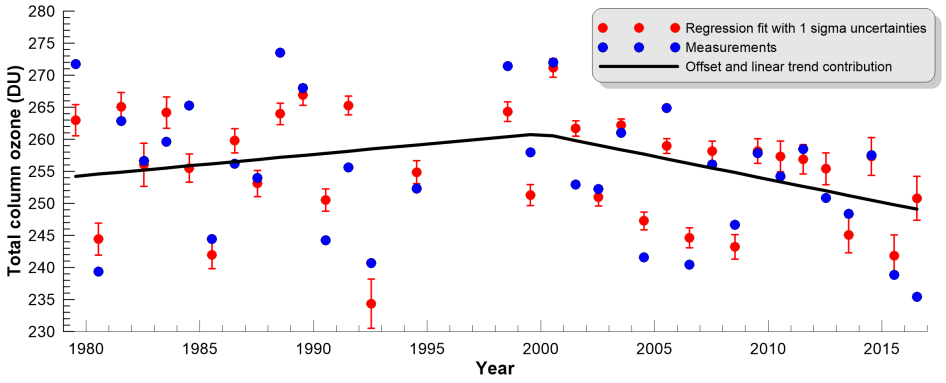

Figure 19. The July mean total column ozone at 19.5° S, 164.375° W (blue dots) together with the regression model values (red dots) and their 1σ uncertainties, and the contribution to the regression model from the offset and two trend basis functions.

Competing interests. The authors declare no competing interest.

Acknowledgements. We acknowledge the World Meteorological Organization-Global Atmosphere Watch (WMO/GAW) Ozone Monitoring Community, World Ozone and Ultraviolet Radiation Data Centre (WOUDC) for the Dobson and Brewer TCO data retrieved 14 August 2018 from http://woudc.org. A list of all contributing sites is available on https://search.datacite.org/works/10.14287/10000001. We would also like to thank the many satellite teams who provide their data freely to the ozone research community. Without access to these source data, the creation of combined TCO databases such as those reported on here would not be possible. The development of the NIWA-BS TCO database was funded under the New Zealand Deep South National Science Challenge while the development of the BS-filled TCO database was funded internally at Bodeker Scientific. We acknowledge that the isolation that comes with the covid-19 lockdown can also, sometimes, have benefits i.e. it provided the head-space for the lead author to get this paper completed and submitted.

References

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.: Continuous decline in lower stratospheric ozone offsets ozone layer recovery, Atmospheric Chemistry and Physics, doi:10.5194/acp-18-1379-2018, 2018

Basher, R.E. and Bojkov, R.D.: WMO-sponsored intercomparisons of Dobson Spectrophotometers improve ozone data quality, WMO Bulletin, 44(3), 248-250, 1995

Bhartia, P. K.: OMI Algorithm Theoretical Basis Document Volume II OMI Ozone Products, ATBD-OMI-02, Version 2.0, 2002, available at: https://docserver.gesdisc.eosdis.nasa.gov/repository/Mission/OMI/3.3_ScienceDataProductDocumentation/3.3.4_ProductGenerationAlgorithm/ATBD-OMI-02.pdf, last accessed 19 April 2021.

Bodeker, G.E., Boyd, I.S., and Matthews, W.A.: Trends and variability in vertical ozone and temperature profiles measured by ozonesondes at Lauder, New Zealand: 1986-1996, J. Geophys. Res., 103, 28661-28681, 1998

Bodeker, G.E., Connor, B.J., Liley, J.B., and Matthews, W.A.: The global mass of ozone: 1978-1998, Geophys. Res. Lett., 28, 2819-2822, 2001

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

Bodeker, G.E., Shiona, H., and Eskes, H.: Indicators of Antarctic ozone depletion, Atmos. Chem. Phys., 5, 2603-2615, 2005

Bodeker, G.E., Hassler, B., Young, P.J., and Portmann, R.W.: A vertically resolved, global, gap-free ozone database for assessing or constraining global climate model simulations, Earth. Syst. Sci. Data., 5, 31-43, 2013

Bodeker, G.E., and Kremser, S.: Techniques for analyses of trends in GRUAN data, Atmos. Meas. Tech., 8, 1673–1684, 2015

Bodeker, G.E., Nitzbon, J., Lewis, J., Schwertheim, A., Tradowsky, J.S. and Kremser, S.: NIWA-BS Total Column Ozone Database, [Data Set], Zenodo, doi:10.5281/zenodo.1346424.

Bodeker, G.E., Kremser, S. and Tradowsky, J.S.: BS Filled Total Column Ozone Database, [Data Set], Zenodo, doi:10.5281/zenodo.3908787.

Bojinski, S., Verstraete, M., Peterson, T.C., Richter, C., Simmons, A., and Zemp, M.: The concept of essential climate variables in support of climate research, applications, and policy, Bull. Amer. Meteor. Soc., 95, 1432-1443, doi:10.1175/BAMS-D-13-00047.1, 2014.

Damadeo, R. P., Zawodny, J. M., Thomason, L. W., and Iyer, N.: SAGE version 7.0 algorithm: application to SAGE II, Atmos. Meas. Tech., 6, 3539-3561, doi:10.5194/amt-6-3539-2013, 2013.

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

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

Frith, S.M., Kramarova, N.A., Stolarski, R.S., Mcpeters, R.D., Bhartia, P.K. and Labow, G.J.: Recent Changes in Column Ozone based on the SBUV Version 8.6 Merged Ozone Dataset, J. Geophys. Res., submitted, 2014.

GCOS-138: Implementation plan for the global observing system for climate in support of the UNFCCC, GOOS-184, GTOS-76, WMO-TD/No. 1523, 2010

GCOS-143: Guideline for the Generation of Datasets and Products Meeting GCOS Requirements, WMO/TD No. 1530, 2010.

Godin, R.: Joint Polar Satellite System (JPSS) OMPS NADIR Total Column Ozone Algorithm Theoretical Basis Document (ATBD), NASA Goddard Space Flight Center, GSFC JPSS CMO, 2014, report available at https://www.star.nesdis.noaa.gov/jpss/documents/ATBD/D0001-M01-S01-005_JPSS_ATBD_OMPS-NP-Ozone_A.pdf, last accessed 3 May 2021.

~~GCOS-143: Guideline for the Generation of Datasets and Products Meeting GCOS Requirements, WMO/TD No. 1530, 2010~~

Hendrick, F., Pommereau, J.-P., Goutail, F., Evans, R.D., Ionov, D., Pazmino, A., Kyrö, E., Held, G., Eriksen, P., Dorokhov, V., Gil, M., and Roozendael, M.V.: NDACC/SAOZ UV-visible total ozone measurements: improved retrieval and comparison with correlative ground-based and satellite observations, Atmos. Chem. Phys., 11, 5975-5995, 2011

Herman, J., Bhartia, P.K., Krueger, A. J., McPeters, R. D., Wellemeyer, C. G., Seftor, C. J., Jaross, G., Schlesinger, B. M., Torres, O., Labow, G. and others: Meteor-3 Total Ozone Mapping Spectrometer (TOMS) Data Products User's Guide, NASA, 1996, available at: https://ozoneaq.gsfc.nasa.gov/media/docs/m3usrguide.pdf, last accessed 19 April 2021.

Krueger, A., Bhartia, P. K., McPeters, R., Herman, J., Wellemeyer, C., Jaross, G.,Cebula, R. and others: ADEOS Total Ozone Mapping Spectrometer(TOMS) Data Products User's Guide. NASA, 1998, available at https://bookshop.org/books/adeos-total-ozone-mapping-spectrometer-toms-data-products-user-s-guide/9781730705618.

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

Lerot, C., Van Roozendael, M., Spurr, R., Loyola, D., Coldewey-Egbers, M., Kochenova, S., van Gent, J., Koukouli, Balis, D., Lambert, J.-C., Granville, J., Zehner, C.: Homogenized total ozone data records from the European sensors GOME/ERS-2, SCIAMACHY/Envisat, and GOME-2/MetOp-A, Journal of Geophysical Research, doi 10.1002/2013JD020831, 2014.

Liddle, A.R.: Information criteria for astrophysical model selection, Mon. Not. R. Astron. Soc., 377, 2007

McPeters, R. D., Bhartia, P. K., Krueger, A. J., Herman, J. R., Schlesinger, B. M., Wellemeyer, C. G., Seftor, C. J., Jaross, G., Taylor, S. L., Swissler, T. and others: Nimbus-7 Total Ozone Mapping Spectrometer (TOMS) Data Products User's Guide, NASA, 1996, available at: https://docserver.gesdisc.eosdis.nasa.gov/public/project/TOMS/NIMBUS7_USERGUIDE.PDF, last accessed 19 April 2021.

McPeters, R. D., Bhartia, P. K., Krueger, A. J., Herman, J. R., Wellemeyer, C. G., Seftor, C. J., Cebula, R. P. and others: Earth probe Total Ozone Mapping Spectrometer (TOMS): Data Products User's Guide, NASA, 1998, available at: https://ozoneaq.gsfc.nasa.gov/media/docs/epusrguide.pdf, last accessed 19 April 2021.

McPeters, R.D., Bhartia, P.K., Haffner, D., Labow, G.J., and Flynn, L.: The version 8.6 SBUV ozone data record: An overview, J. Geophys. Res., 118, doi:10.1002/jgrd.50597, 2013

Müller, R., Grooß, J.-U., Lemmen, C., Heinze, D., Dameris, M., and Bodeker, G.E.: Simple measures of ozone depletion in the polar stratosphere, Atmos. Chem. Phys., 8, 251-264, 2008

Newman, P.A., Daniel, J.S., Waugh, D.W., and Nash, E.R.: A new formulation of equivalent effective stratospheric chlorine (EESC), Atmospheric Chemistry and Physics, 7, 4537–4552, 2007

Press, W.H., Flannery, B.R., Teukolsky, S.A., and Vettering, W.T.: Numerical Recipes in Pascal, 759 pp., Cambridge Univ. Press, New York, 1989.

Saha, S., Moorthi, S., Pan, H., Wu, X., Wang, J., Nadiga, S., Tripp, P.c., Kistler, R., En, J.W., Behringer, D.d., Liu, H., Stokes, D., Grumbine, R., Gayno, G., Wang, J., Hou, Y., Chuang, H., Juang, H.H., Sela, J., Iredell, M., Treadon, R., Kleist, D., T, P.V.D., Keyser, D., Derber, J., Ek, M., Meng, J.e., Wei, H., Yang, R.a., Lord, S., van den Dool, H., Kumar, A., Wang, W.u., Long, C., Iah, M.C., Xue, Y., Huang, B.n., Schemm, J., Ebisuzaki, W.e., Lin, R., Xie, P., Chen, M., Zhou, S., Higgins, W., Zou, C., Liu, Q., Chen, Y., Han, Y., Cucurull, L., Reynolds,

R.W., Rutledge, G., and Goldberg, M.: The NCEP Climate Forecast System Reanalysis, Bulletin of the American Meteorological Society, 1015-1057, 2010

Schwarz, G.: Estimating the dimension of a model, The Annals of Statistics, 6, 461-464, 1978

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

Weber, M., Chehade, W., Fioletov, V.E., Frith, S.M., Long, C.S., Steinbrecht, W., and Wild, J.D.: Stratospheric Ozone in [Blunden, J., and D.S. Arndt, State of the Climate in 2012]. Bull. Amer. Meteor. Soc., 94, doi: 10.1175/2013BAMSStateoftheClimate.1, 2013.

Wild, J.D., Long, C.S., Barthia, P.K., and McPeters, R.: Constructing a Long-Term Ozone Climate Data Set (1979-2010) from V8.6 SBUV/2 Profiles, Quadrennial Ozone Symposium 2012, Toronto, Canada, 2012.