Response to referee comments

May 15, 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 which 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 overestimated 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 in any biases with respect to ground-based measurements - 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 below. 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

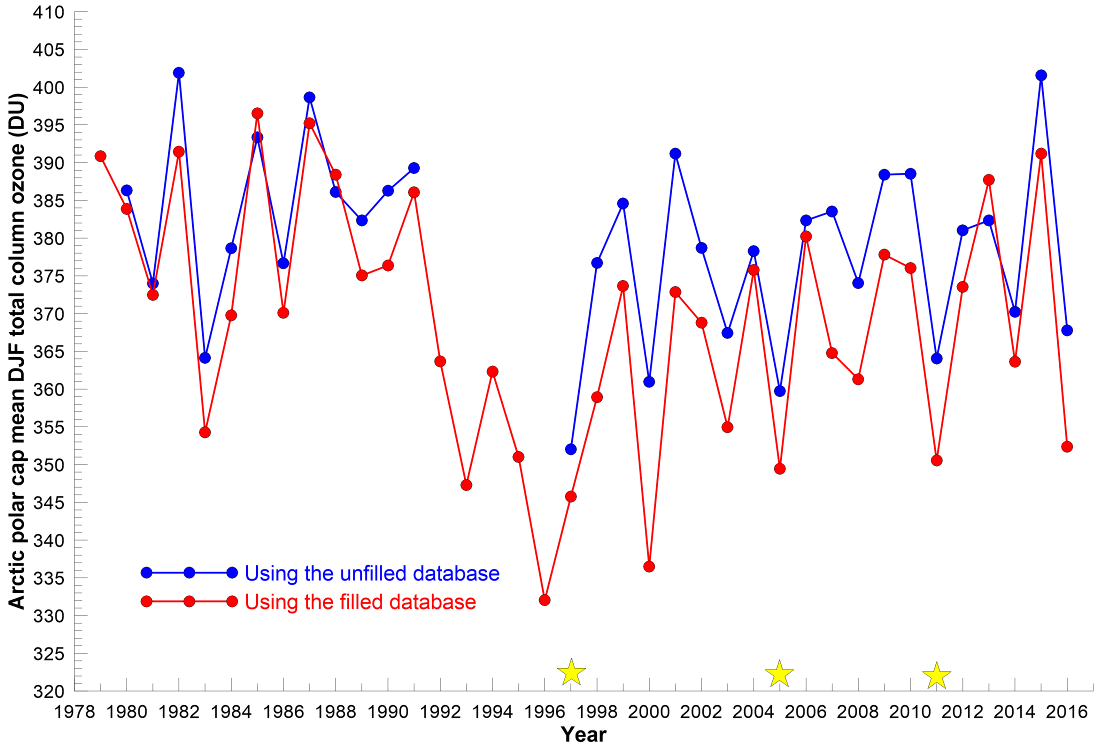

Figure 1: Arctic polar cap area weighted mean TCO for DJF calculated from the filled (red) and unfilled (blue) database. Stars show the year highlighted by the reviewer.

algorithm performs in the case of minor gaps.

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 width 10, 20, 30, 60 and 120° (see Figure 3). The data filling algorithm was then applied to the field shown 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:

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

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.

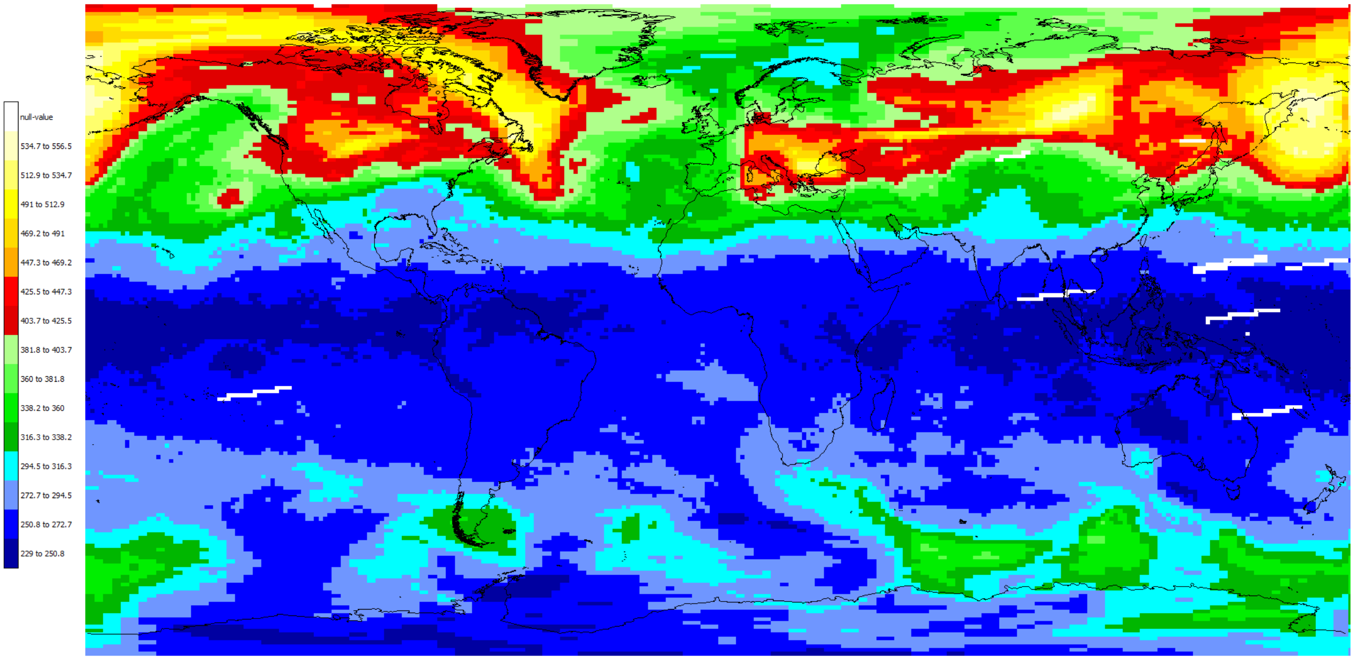

Figure 2: The original TCO field on day 80 of 1982.

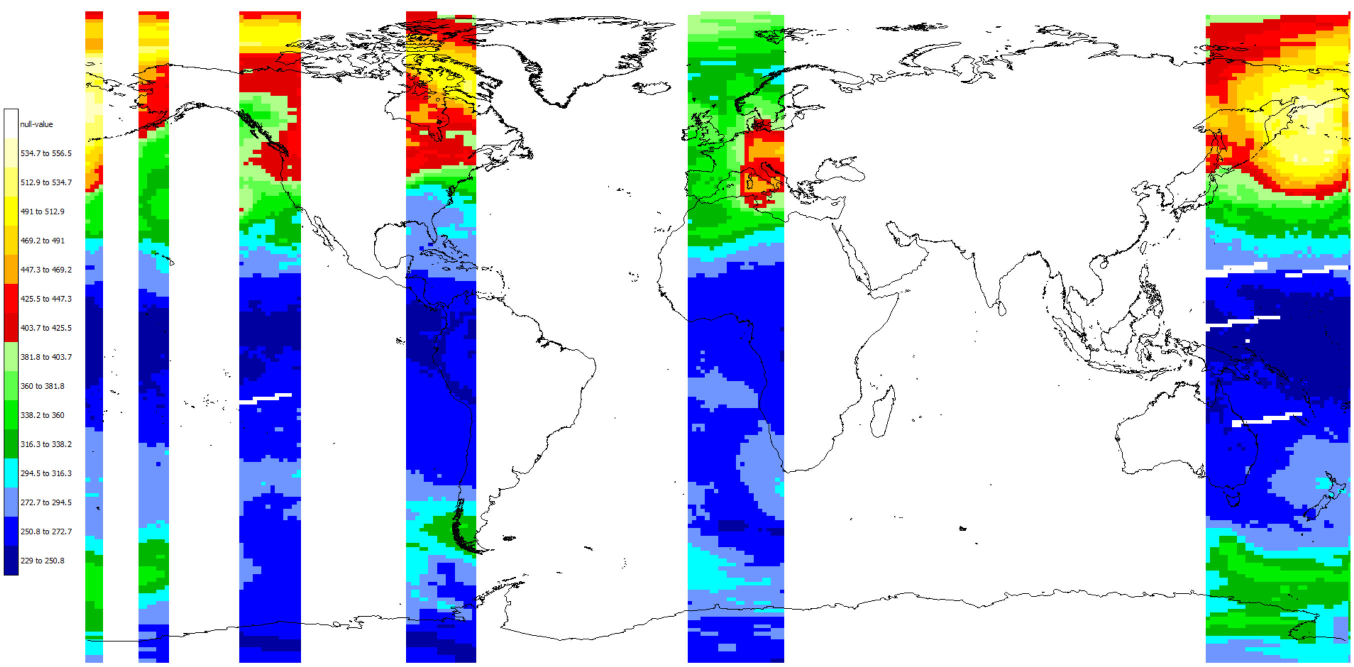

Figure 3: The TCO field on day 80 of 1982 with swaths of width 10, 20, 30, 60 and 120° having been removed.

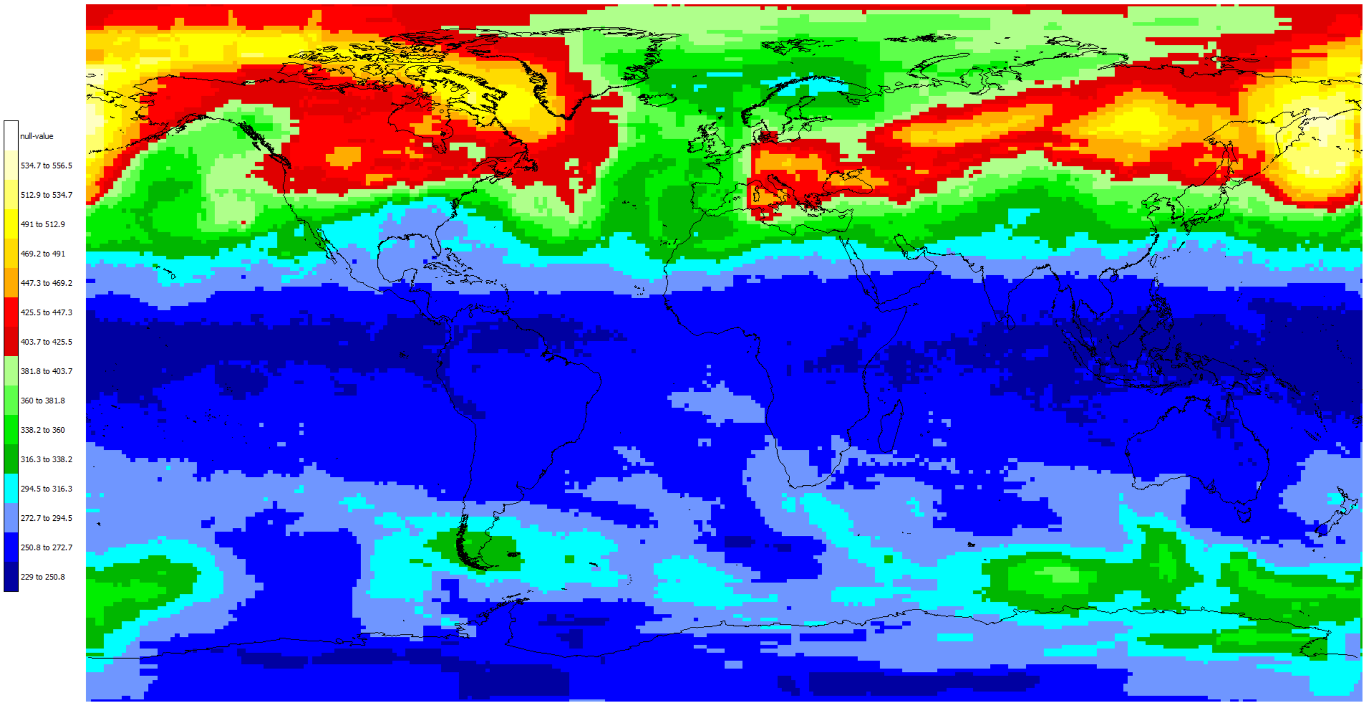

Figure 4: The filled TCO field on day 80 of 1982.

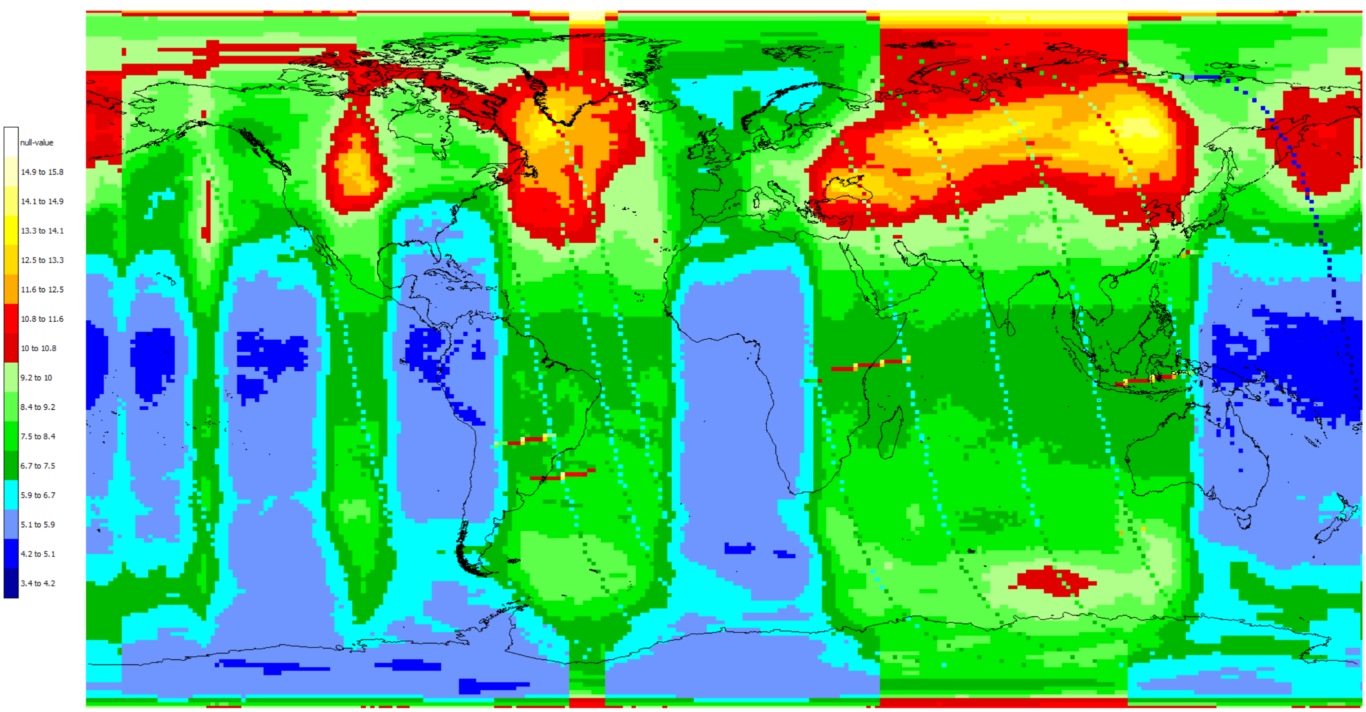

Figure 5: The uncertainties on the filled TCO field for day 80 of 1982.

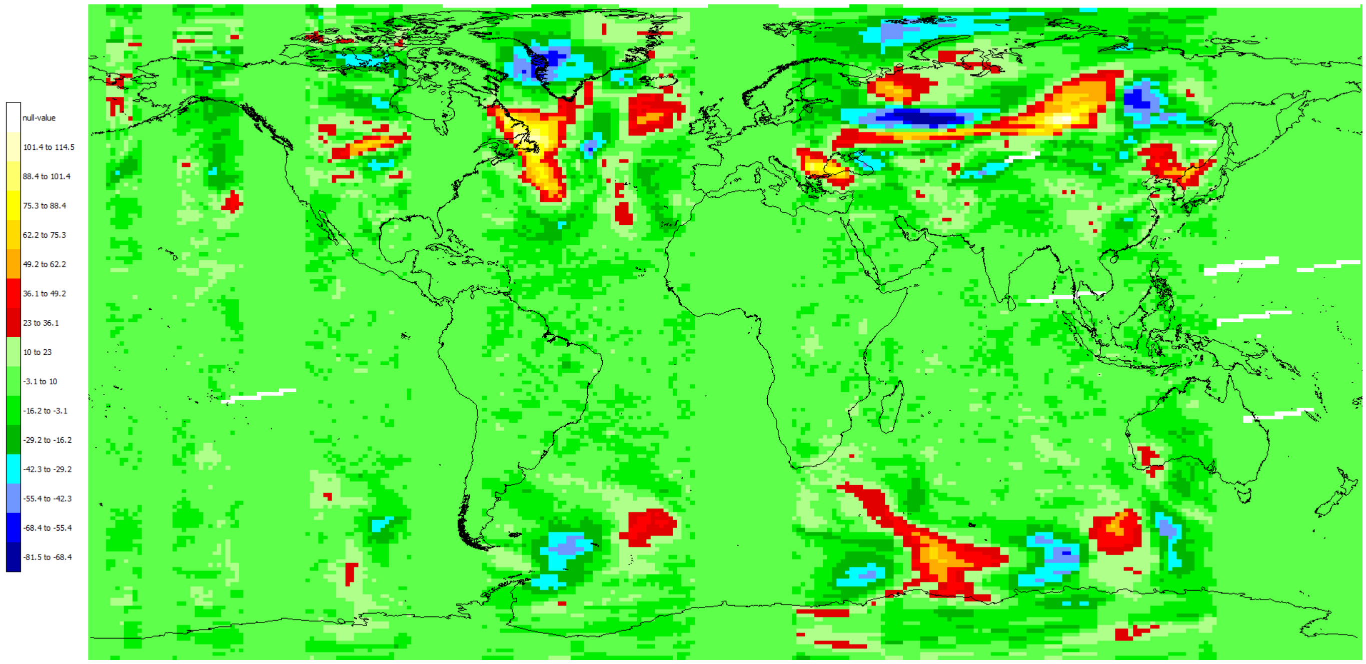

Figure 6: The differences between the original TCO field shown in Figure 2 and the filled field shown in Figure 4.

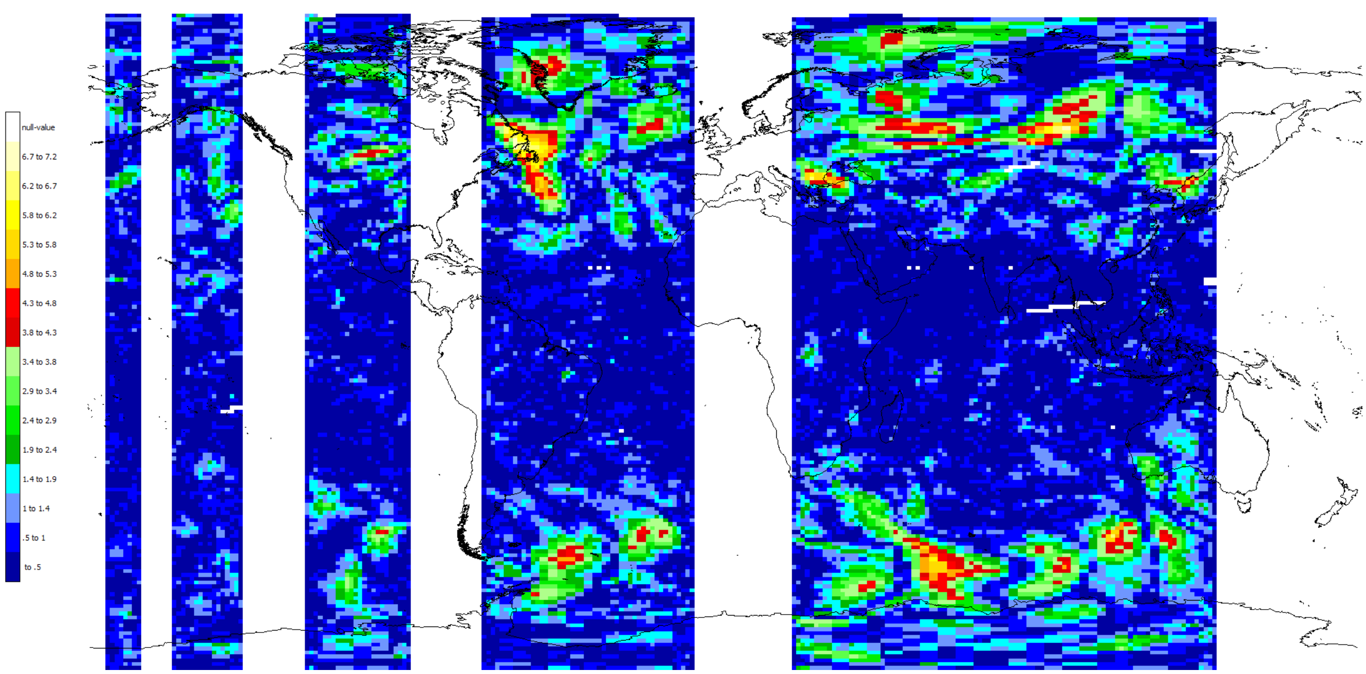

Figure 7: The k values calculated for the filled field. Only k values for the filled values are shown since they are otherwise everyone 0.0 since there $m_1 = m_2$.

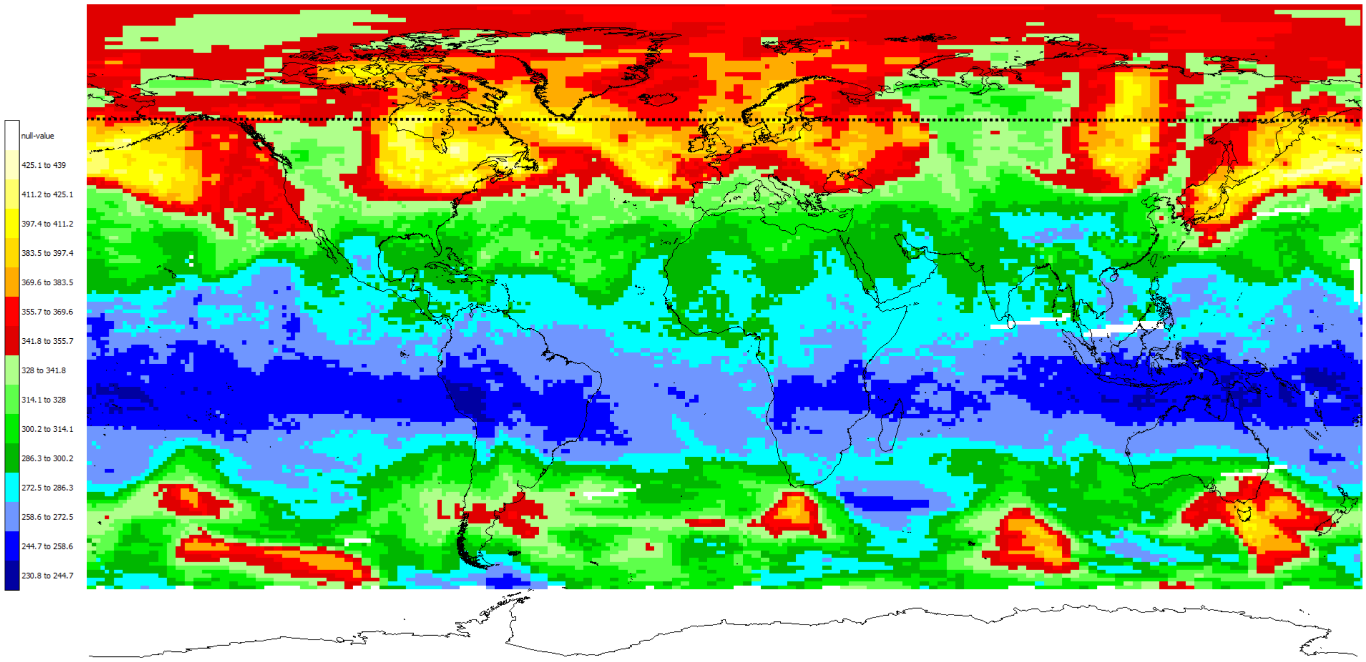

Figure 8: The original TCO field on 21 June 1982. Data north of 60°N were deleted (i.e. poleward of the dashed line).

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. in 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 Principle Investigators of the space-based instruments (in particular Rich McPeters, PK Bhartia and John Burrows)

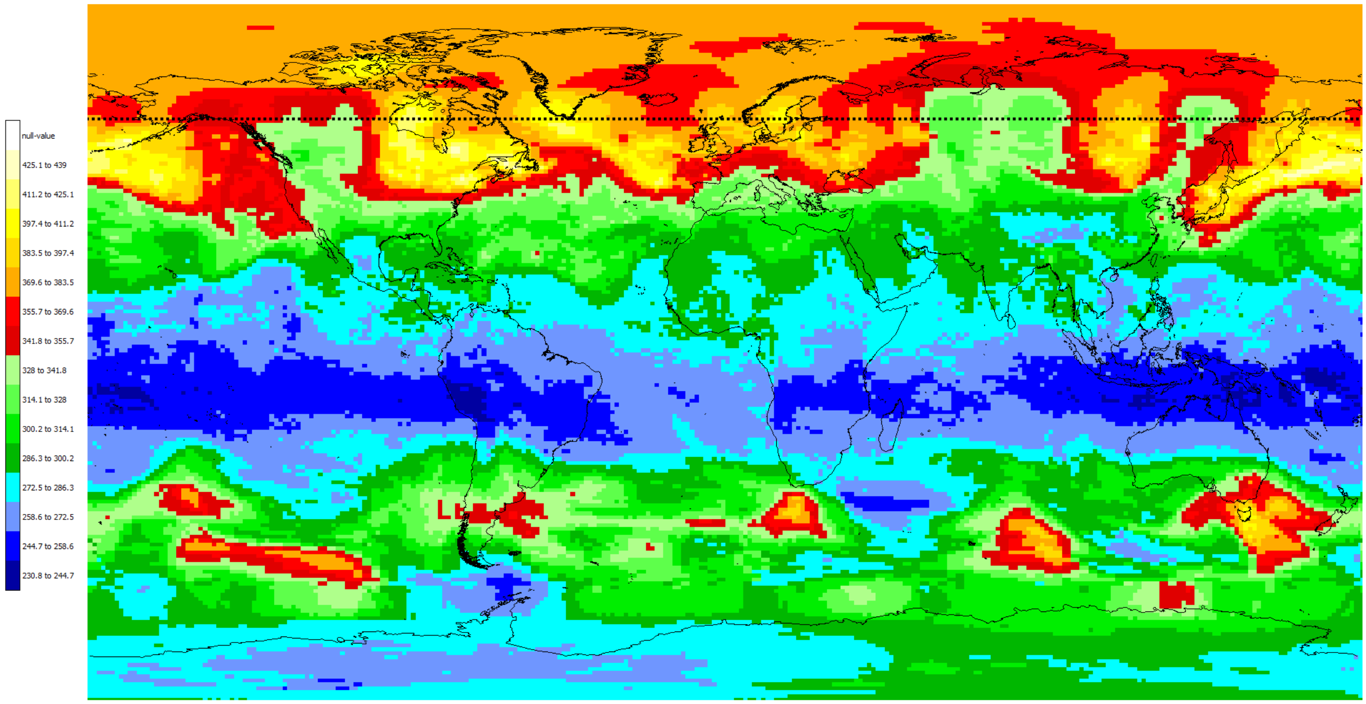

Figure 9: The filled TCO field on 21 June 1982. Data poleward of the dotted line (60°N) were filled.

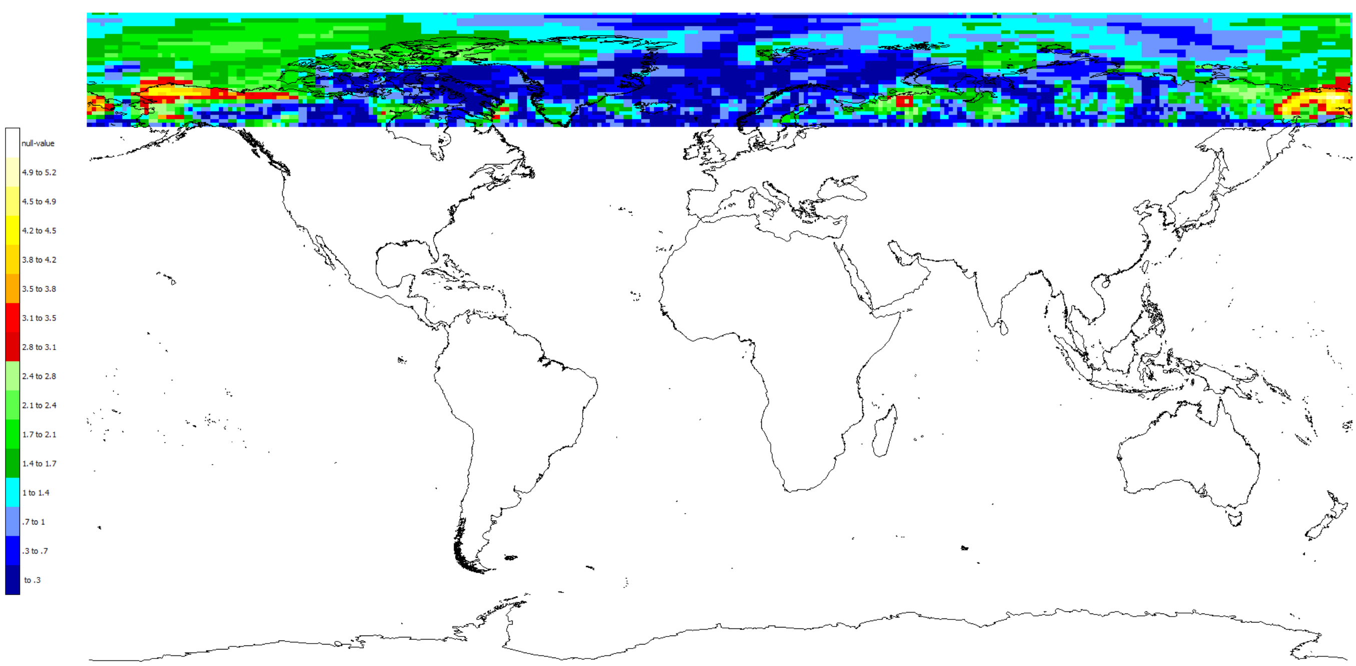

Figure 10: The k values calculated for the Arctic filled data.

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 versions of this database, 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 three 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.

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

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.