Earth Syst. Sci. Data Discuss.,
https://doi.org/10.5194/essd-2020-218-AC3, 2021

Earth System Science Data
Open Access Discussions

ESSDD

Interactive comment

Interactive comment on "A Global Total Column Ozone Climate Data Record" *by* Greg E. Bodeker et al.

Greg E. Bodeker et al.

greg@bodekerscientific.com

Received and published: 15 May 2021

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:

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.

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 $1.25^\circ \times 1.0^\circ$ 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.

Interactive comment on Earth Syst. Sci. Data Discuss., https://doi.org/10.5194/essd-2020-218, 2020.