Replies to the Editor:

Comments to the authors

Thank you for the thorough revision of your manuscript. I am happy to accept this paper if you address the two following minor issues:

1) "Stippling" in the figures, and in particular Figure 3: first, please abandon the term "statistical significance" (see https://www.nature.com/articles/d41586-019-00857-9 and the more detailed articles in the literature). The implication of "significance" and of the stippling in the plots is that the results are not meaningful if they are "not significant", but this is incorrect. I am not against stippling per se, although I much prefer to use two different saturation levels as this is psychologically much more intuitive and does not imply that you wish to discard stippled features. Please take time to think this through thoroughly, because it is very important for all scientific work.

   ## We are thankful to the Editor for raising this important aspect of manuscript writing. We agree with the suggestion and in the further revised manuscript we no longer use the term "statistically significant" anywhere. Now captions for Figures 3 and 6 clearly state that "stippling indicates the regions where differences are smaller than 1-sigma variability".

2) I would like to see a better justification for including the tropospheric levels in the data product, even if they are almost exclusively TOMCAT simulation based. Please look at this from a user perspective: how do you envision your data product to be used? Is it for example important that the total ozone column is meaningful? (This means that stitching together ozone fields from two different product typically needs some bias correction). I think it is good to have the whole atmosphere covered, but I would like to be convinced more strongly.

   ## We agree with the Editor that some better explanation was needed for the including ozone profiles at all the levels. In the Methodology section we have clarified that the TOMCAT tropospheric values are based on a climatology (Logan et al., 1999) in the stratospheric model, and that these values are then used unchanged in ML-TOMCAT below 316 hPa. Updated files are uploaded on Zenodo and both the versions can be accessed via: https://zenodo.org/record/5651194

   We added an additional section (Section 4.3.4) to discuss total column comparisons and show that having whole atmosphere profiles helps to improve total column calculation. We have now added an additional figure in the manuscript (Figure 10) comparing total column ozone for six latitude bands (Arctic (60N-90N), Antarctic (60S-90S), NH mid-latitudes (35N-60N), SH mid-latitudes (35S-60S), tropics (20S-20N) and near global (60S-60N)) with SBUV merged data. The comparison clearly shows that ML-TOMCAT total column data shows much better agreement with the SBUV data compared to TOMCAT data. We make it clear that we include the tropospheric values to allow column comparisons such as this, but the tropospheric values should not be used for tropospheric studies. We also added a supplementary figure (Figure S22) to compare tropospheric ozone columns from ML-TOMCAT and TOMCAT. The TOMCAT data set shows a near-identical annually repeating cycle whereas tropospheric ML-TOMCAT shows variability with some short-term and long-term variations at different latitude bands. A further figure (S23) compares the TOMCAT and ML-TOMCAT mean latitude-altitude ozone cross section.