

Review on "**Improved CCD tropospheric ozone from S5P TROPOMI satellite data using local cloud fields**" by Swathi Maratt Satheesan, Kai-Uwe Eichmann, John P. Burrows, Mark Weber, Ryan Stauffer, Anne Thompson, and Debra Kollonige

General comment:

The manuscript "Improved CCD tropospheric ozone from S5P TROPOMI satellite data using local cloud fields" by S. Satheesan describes improved versions the traditional CCD approach to calculate tropospheric ozone. Three versions are presented and compared to sondes data from the SHADOZ network and the operational S5P tropospheric ozone column data product:

- 1) CPC uses the traditional approach and differs to the operational CCD data mainly in a newer climatology for ozone mixing ratio around cloud top height.
- 2) CLC exchanges the fixed pacific reference sector by a local stratospheric reference.
- 3) CLCT replaces the used climatology by a linear regression (Thiel-Sen) to interpolate or extrapolate to the fixed reference pressure level and uses the local stratospheric reference.

The major advantage of the local reference algorithms is that the assumption of a longitudinal invariant stratospheric column is no longer needed. This offers the possibility to extend the retrieval from the inner tropical band ($\pm 20^\circ$) also to the subtropics (25°N to 25°S).

The authors show that the local reference algorithms compare better with the SHADOZ sonde data than the fixed reference datasets.

major comments:

As far as I understood the CHORA data are provided on a daily basis. To me it is not clear whether the validation is performed on a daily or monthly basis:

1127: "*Ozonesonde profiles are considered collocated to the satellite data when the station is inside the spatio-temporal grid-box of the TROPOMI tropospheric ozone data, here $0.5^\circ \times 0.5^\circ$ and the same day*". (daily comparison)

1246: "*The monthly mean of tropospheric ozone retrievals ... from all three CHORA algorithms were calculated.*" (sounds like monthly data)

Despite the first statement it seems to be on a monthly average: To a large extend the deviations in the subtropics are driven by the March soundings in Hanoi. As one explanation the authors mention "*The challenge of limited ozonesonde measurements in March further complicates the validation.*" (L274) You would need at least 15-20 measurements per month to represent the monthly mean values adequately. This is not given at any sounding station. Because of that I recommend comparison on a daily basis.

The description of the CPC algorithm (fixed pacific reference) states that it is designed for the 20°S to 20°N latitude band. However through out the manuscript it is applied from 25°S to 25°N and beyond e.g. figure 8 (comparison to Irene 25.9°S). When looking at the operational CCD dataset, the data in the "winter hemisphere" are often corrupted due to the lack of convective clouds outside the 15° range. Here some more details should be added how the respective extension is achieved.

minor comments:

l 124: "and when the burst height of the sonde is below 15 km." replace the "and" by "or"? what is the reason for the second requirement? Of course, if the burst height is below the 270 hPa altitude level the data are not useful but it does not matter the balloon bursts at 14.8 km

p 5: The authors should decide whether the column between the surface and 270hPa is above (caption figure1) or below (l 134) this pressure levels. From my point of view both have their justification: "above" (because the pressure gets higher) or "below" as the altitude is below the 270hPa level, but using both (within a few lines) is a bit confusing.

l 151: The product readme file recommends a qa-value greater than 70 (reference ESA_21b, https://sentinel.esa.int/documents/d/sentinel/s5p-mpc-dlr-prf-o3_tcl_v02-06-01_2-7_20231129_signed). This will probably not improve the validation with sondes.

p5 and 6 I suggest to combine the descriptions of the operational ESA product and the CPC. The algorithm is basically the same except for the climatology and the daily vs 3 day averages.

l 165 The CPC algorithm is described to be applied between 20°S and 20°N but in figure 13 data up to 25° South and North are shown (see also major comments).

l 216: This sentence is unclear: " ... by substituting the values of slope and medians of the ACCO ..." by what? and why replacing the median of the ACCO ?

figure 6: "latitude band 2.4° S - 0.4° S and longitude band 26.8° E - 46.8° E" according to the previous description I expected a latitude range of $\pm 1^\circ$ for a grid box of 0.5°, so 1.25°S +1° = 0.25°S to 2.25° S but not 2.4°. The same holds for the longitudinal range. I assumed the centre of the grid box to be the respective starting point.

How fast are the longitudinal ranges extended. Starting with $\pm 5^\circ$, what is the next step if the number of cloudy data is too low, $\pm 10^\circ$?

l 245 I suggest to look for additional sonde data at <https://woudc.org/data/explore.php?lang=en> especially for the subtropical areas (Hong Kong "King's Park")

l 271 as well as in many other places (figures 10 and 11) throughout the manuscript, the seasons are reversed to as Spring to Winter, according to our northern hemispheric seasons. I am not sure if this useful in the tropical to subtropical regions. I suggest to use the month instead, or local dry and wet season.

figures 8 and 9 Can you include the operational O3_TCL product in this figures? It is included in figure 13 so the analysis has already been done anyway.

figure 10: this is a very interesting figure - maybe you could make more use of it? For example the median distance of deep convective clouds over Ascension Island is (3000km) more or less independent of the season. That means the deep convective clouds are over the south American continent.

figure 11: include the 270 hPa level in the figure

Discussion of individual sounding stations.

For some stations it might be nice to have figure of the differences per month, however including this in the manuscript might extend it too much. How about adding such figures to a supplement?

l 315ff (Paramaribo) How about local sources, biomass burning might also occur in the vicinity of Paramaribo?

l 343-346 (Kuala Lumpur) The CPC compares better to the sondes than the local cloud reference algorithms. Does this imply that for the local reference algorithms the reference area might be too small and should be increased by default. How large are the local reference areas for a typical dry and wet season? According to figure 10 the mean distance is very small (~ 200 km) only for the Nov-January region the distance is slightly larger.

figure 13: may be a longitudinal plot might be more telling. The fixed reference region algorithms (CCD-ESA and CPC) compare best to the sondes in the reference region (70°E to 170°W, Fiji, Samoa and Kuala Lumpur), outside the reference region the CLC algorithms might show their capabilities.

l 481: this sensitivity study might be interesting, parts of it might be shown in a supplement

l 500 shift figure 13 to here.

technical comments:

l 58 "... in the time frame of 2017 -2022" even if this was the original plan, meanwhile it's 2024 S5P is still operational and we get very nice results every day. I suggest to replace by "...the time frame after 2018"

l 84/85 "3.5 x 7.5 km² (across-track x along-track), was further refined to 5.5 x 3.5 km²". exchange the updated resolution to 3.5 x 5.5 km² to have across-track x along-track.

l 475 "significantly small", what does this mean? significant or small or significant but small?