Anonymous Referee #3

Received and published: 9 March 2016

The paper presents measurements of tropical tropospheric ozone columns from nadir retrievals of GOME, SCIAMACHY and GOME-2 using the CCD technique. The paper is well structured and written and in principle can be accepted for publication after considering my comments below.

Page 3, lines77-88. What is the difference between the Valks et al., retrievals and the current ones, apart from including SCIAMACHY? Why the authors don't compare these later on or at least provide some relevant comments?

Both algprithms are quite similar. One of the main differences are the corrections applied to ACCO (climatological correction, screening out outliers, correction for ozone overestimation in the case of FRESCO), the grid size used and the use of different criteria for selecting ACCOs depending on the instrument used.

As it is pointed out in the introduction:

"The CCD_IUP algorithm has been initially developed as a verification algorithm for S5P_TROPOZ_CCD prototype algorithm within the framework of the TROPOMI algorithm development which is based upon the Valks et al. algorithm (Richter A. and the Verification Teams, 2015). The scope of the verification is to identify possible deficiencies, and highlight potential improvements. Both algorithms were applied to operational GOME-2 GDP 4.7 data. The agreement between them was found to be very good (0.9 < R < 0.99, RMS between 4 and 9 DU and the biases less than 2 DU in most of the cases)."

Page 4, line 108-117. To my understanding the application of WF-DOAS to each instrument is treated with a different cloud- algorithm. How this could affect the retrievals when applying the CCD technique?

In order to retrieve ACCO for the three different satellite instruments we use different cloud selection criteria adapted to the peculiarity of each cloud algorithm in order to have a statistically sufficient sample of cloudy data (>1% monthly averaged ozone data per latitude band) as a basic condition.

As referred in the manuscript:

"In order to define the DCCs we used measurements with cloud fractions (cf) greater than 0.8 and cloud top heights (cth) greater than 9 km for SCIAMACHY and 7 km for GOME and GOME-2. Even with a higher cth threshold for SCIAMACHY, SCIAMACHY has the highest frequency of "cloudy" measurements among the three satellites. Fig. 3b shows that roughly ~25% of cloud top heights in the western Pacific are higher than 9 km for SACURA (SCIAMACHY), whereas for FRESCO (GOME-2), the same frequency is met for clouds only above 7 km. Since the cloud algorithms differ between instruments and in order to have sufficient "cloudy" ozone measurements in more than 1% of all the measurements per latitude band, the lower cloud top height limit classifying the DCCs is different for each satellite instrument."

Page 5, line 135-137. Are these thresholds consistent or independent from the cloud algorithm applied?

For all cloud algorithms i the variation of the ACCO for different critical values were tested. We concluded that the ACCO does not change significantly using clouds higher than 7 km and cloud factions greater than 0.8. The standard deviations are greater for latitude bands with less than 1% cloudy data. Due to the ITCZ nortward and southward movements northwards less cloudy data are available in the respective winter hemispheres. The parameter changes dramaticallywhen the number of cloudy measurements drop dramatically (see figure 3b) in cases of very high cloud top height and cloud fraction limits.

This is described as follows in the manuscript:

"Since the cloud algorithms differ between instruments and in order to have sufficient "cloudy" ozone measurements in more than 1% of all the measurements per latitude band, the lower cloud top height limit classifying the DCCs is different for each satellite instrument.

It is obvious that the different cloud algorithms calculate different cloud fractions and top heights and as a result yield different ACCO values (see Fig. 4a). However, it was concluded that the ACCO does not change significantly when the cf is greater than 0.8 and cth greater than 7 km. For the calculation of the ACCO, all "cloudy " measurements (defined separately for each instrument in order to have enough data) are selected and monthly averaged in latitude bands of width 2.5◦ between -20◦S and 20◦N in the western Pacific and Indian Ocean (70◦E-170◦W)."

Page 5, line 148-149. How do the authors explain the difference in frequency of cloudy measurements for SCIAMACHY?

The differences between FRESCO and other cloud algorithms are well described in:

Lelli L.: Studies of global cloud field using measurements of GOME, SCIAMACHY and GOME-2, PhD Thesis, , University of Bremen, 2013.

It is planned to use a common cloud algorithm for all instruments in the future.

Page 5, line 163-165. Why the authors choose as a fixed level approximately 12km. Is this not low as an indicative tropopause height for the tropics? On which data/model is it based the climatological correction term?

This level may be low for the tropics but on the other hand the highest DCCs retrieved with FRESCO (as used for GOME and GOME-2) algorithm reach only 14 km. In order to have valid statistics, we had to drop the level. Another reason why we use the 200 hPa level is because we want to avoid as much as possible the stratospheric intrusions at the tropical borders.

Page 6, 188. What does “50 cloudy ozone measurements per latitude band” represent? Is it arbitrary?

This critical value has changed to 1% of the total measurements that have to be classified as DCCs per latitude band.

Page 6, 191-193. Any justification for the limits of 10DU and 4DU?

The 10 DU is approximately the 2 sigma standard deviation of the ACCO and the 4 DU have been changed to 5DU which is approximately 1 sigma.

Page 6, line 200. Correct to "above 200hPa at".

It is changed to:

" The comparison of the ozone column above 200 hPa with six ozonesonde stations "

Page 7, 203. How many data were available per month from the ozonesondes and how comparable is this number with the ACCO estimates used for averaging?

The number of ozonesonde data for this month varies between 1 and 4 per station while the ACCO measurements per latitude band are on the order of hundreds.

Page 7, line209-211. Figure 4b and its explanation is hard to follow. What we see apart from the overestimation is also a hemispherical difference. This figure will need more discussion.

This figure has been removed since it does not provide any valuable information explaining the method.

Section 4. This section is not well written and confuses the reader how the author use the term error (random, systematic) and how the error terms have been identified and propagated.

The whole section has been re-written. Section 4, provides an uncertainty estimation including the standard uncertainty of the retrieved TCO and ACCO as well as taking their means and the uncertainties arising in the ACCO from the propagated uncertainties in the cloud fraction and cloud top height determination.

Section 5.1. The authors should comment why GOME-2 shows a better correlation than the other two sensors when compared to the ozonesondes. They should also comment why the difference between TTCO and ozonesondes also shows some seasonality especially in Ascension. In general this section mixes two different issues and confuses the reader. The authors discuss in parallel differences in tropospheric ozone between different regions and validation-comparison results. I would suggest to separate the discussion.

In the section "validation with ozonesondes", the second paragraph discusses the results of the validation and the following paragraphs are discussing the time series range and the particularities of each region.

Limiting the comparison to the grid-box where the sonde belongs to, there is no clear seasonality in the TTCO differences in Ascension.

The following text comments on the validation results for each instrument:

" Comparing the CCD_IUP results for each instrument with ozonesondes, we conclude that SCIAMACHY TTCOs have the smallest relative differences (8–19%), bias (-3.7 DU), and RMS (2.6–8.4 DU) with the ozonesondes. Additionally, in more than half of the stations (five over nine) there is a

strong correlation (R> 0.6). GOME CCD_IUP results on the other hand, have generally small biases with the ozonesondes with the exception of Hilo, where the bias is 4.1 DU. The RMS differences are less than 7 DU (with the exception of Hilo, where the RMS is up to 10 DU) and the correlation is strong only at three over nine stations used for the comparison. GOME-2 CCD_IUP results compared with the nine ozonesonde stations present consistent positive bias (2–5.6 DU). The RMS differences are greater (4.8–9.2 DU) among the three instruments and the correlation is strong at four stations."

Section 5.2 This sections is poorly discussed in the context of the current paper and could raise more questions than provide evidence for the quality of TCCO using the CCD technique. I would suggest to omit this from the paper.

This section has been re-written and the comparison with the sondes has been removed. The two datasets are now compared directly with each other in Figs. 15 and 16.

In order to make the LNM columns comparable to the CCD TTCOs, we adjust the LNM columns to the 200 hPa level using climatological values from LNM for the Fortuin et al. (1998) climatology. Therefore, we subtract the ozone between the tropopause and the 200 hPa. The LMN data have been gridded with the same grid used for CCD.