

Answers on Anonymous Referee #1
Received and published: 7 November 2017

General Comments:

This study attempts to harmonize a 20-year satellite record including GOME/ERS-2, SCIAMACHY/Envisat, and GOME-2/MetOp-A using SCIAMACHY as a transfer standard. The authors use 6 different schemes in their attempt to harmonize the data and evaluate the relative success of the various approaches via comparisons with in situ measurements (i.e., sondes) when and where possible. The authors suggest that by using 6 different approaches, they are better able to estimate the uncertainty in apparent trends owed to the harmonization itself. Like prior studies, the authors find few areas of significant tropospheric ozone trends. Their analysis of trends over tropical mega cities seems to produce results not too different from prior studies. Frankly, I am really not sure what to take away from this study, and I have several important issues with what is (and is not) presented.

The authors should be aware of a major reprocessing effort of the Southern Hemisphere Additional Ozonesonde (SHADOZ) network data being led by Thompson and Witte. While I realize the papers on their work are just making their way into the literature, their efforts have been ongoing for several years. Since the authors of this study leverage the SHADOZ data, I am surprised that the paper communicates no awareness of the reprocessing effort nor of its potential impact on the results of this study.

At the very least, the authors could have contacted SHADOZ PI Thompson to make sure she was aware of this study and had the opportunity to communicate important updates relevant to the authors. While entirely up to the authors, having Thompson as a co-author would have strengthened the credibility of the sonde results presented in this paper.

Finally, while the paper appears to present different approach to tropical tropospheric ozone trend analyses that recent studies, I do not find the results particularly compelling or worthy of publication in this form at this time.

Recommendation:

I recommend this manuscript be declined for publication in ACP at this time pending major revisions.

Our answer:

Many thanks for the very helpful comments. As proposed by the reviewer, we have contacted Drs. Anne Thompson and Bryan Johnson who are responsible for reprocessing the ozonesondes data from SHADOZ network. Both accepted our invitation to become co-authors of the paper and supported us in revising the paper. We now used reprocessed SHADOZ data (available for American Samoa and Paramaribo). The recent reprocessing mainly focused on improving old sonde data at the stratospheric ozone peak and above. It had very minor impact on our findings. The stations of Hilo and Fiji have been removed from the comparisons. Fiji is affected by air masses originating from the

mid-latitudes and the upper troposphere and Hilo is strongly affected by volcanic outgassing, resulting in negligible ozone concentrations in the boundary layer.

Additionally, Klaus Peter Heue is included as co-author in the paper for granting us access to his tropospheric ozone data which we used for some comparisons.

Since Reviewer #2 suggested to shorten the discussion on the various merging approaches we moved some material into a supplement and discuss mainly the lessons learned from looking at trends from differently merged datasets in Section 2. The estimation of the mean tropical trends is moved to Subsection 4.1 and is now limited between 15°S and 15°N since the tropical borders are strongly influenced by air masses being transported from the mid-latitudes and stratospheric intrusions (Thompson et al., 2017).

Section 3.3.3. about trends in megacities has been removed from the revised version of our paper as also suggested by reviewer #2.

At various places we have expanded on the comparisons to the Heue et al. results. Although similar instruments have been used, the results from this study and Heue et al. are different and are discussed in more detail (see detailed comments).

Our main findings can be summarized as follows:

Tropical tropospheric ozone trends critically depend on the merging/harmonisation approach. This was investigated by investigating six different merging scenarios. The trend of tropical tropospheric ozone is estimated using a multiple linear regression model and for all six scenarios the sensitivity of the derived trends to the harmonisation approach is investigated. Such an approach has not been reported before and may explain why tropospheric ozone trends from different studies do not agree (see e.g. TOAR report). The main conclusion is that the (statistical) trend uncertainties from one scenario may be smaller than the variation of trends from the different merging approaches, which means that the trend uncertainties are in reality larger. At the end we selected the preferred merging scenario by comparing these six merged datasets with ozonesonde data (some of them reprocessed now) from the SHADOZ network.

Detailed Comments:

Page 1

--Line 3: What does “good agreement” mean? Quantify.

We consider the bias between the CCD retrievals and the integrated O₃ profiles from ozone sondes good since they are less than 6 DU which is about the 1sigma uncertainty of the mean station bias (RMS in Table 2 of Leventidou et al., 2016).

We changed in the main text (page 4, line 30) as follows: “The biases between them have been found to be within 6 DU which is mostly within the uncertainties of the mean biases of 6 DU (1 sigma). One large source of uncertainties in these comparisons are the

low sampling of the sondes (less than five launches in a month typically) and the fact that CCD ozone is only derived as monthly means covering rather large areas (grid boxes). In the abstract we only mention the average bias between CCD and sondes”.

Line 14: “Additionally, over central ...” Awkward sentence.

The sentence has been changed to “... and by ~2DU/decade over central ...”

Line 19: “... reasons for these decreases are...”

The sentence has been removed.

Page 2

--Line 2: delete “both”

Deleted

Page 3

--Line 10: “3.8% decade⁻¹ (0.16 ppbv year⁻¹)” What is the difference between these numbers? Unclear. You mention “surface and ozonesonde observations,” so which is it?

The trend in ppbv year⁻¹ represents the change of surface ozone in volume mixing ratio per year. The sentence (page 3, line 5) has been changed to: “Oltmans et al. (2013) observed an increase of 3.8 % decade⁻¹ (0.16 ppbv year⁻¹) in surface ozone in Mauna Loa, Hawaii (19.5°N) in the North Pacific since 1974 and a smaller insignificant trend in the order of 0.7 % decade⁻¹ (0.01 ppbv year⁻¹) in American Samoa (14.5°S) after 1976. ”

Page 4

--Line 4: The reference to Leventidou et al., 2016 may be a recent one for the CCD method, but I think you should be referencing the original paper for this approach, which I believe goes back to Ziemke ..

As the referee mentions, the CCD method was developed by Ziemke et al., 1998 and further improved by Valks et al., 2003. The citation of Ziemke et al. 1998 has been added in the introduction, along with the most significant contributors on tropospheric ozone retrievals from remote sensing in the past.

--Line 10ff: This would be a good place to remind the reader of the specific application of the CCD approach you’re using. To what altitude is tropospheric column ozone being computed? Does it vary scene-to-scene? On what fraction of pixels can it be applied?

The CCD method is described in detail in Leventidou et al., 2016. All tropospheric O₃ columns are calculated up to 200 hPa. The main reason is that most clouds do not reach the tropopause.

--Line 14: “overpass time” – is this not a critical element influencing tropospheric ozone, especially in regions near megacities?

For tropospheric trace gases that show diurnal variations, overpass time is important. All satellite data used here are in the morning hours differing at most one hour (9:30 to 10:30). This is believed to have little impact on the results.

The following text has been added (page 9, line8): “As seen in Table 1, the mean bias between the six harmonised TCO datasets and the ozone sondes range between -1.1 and 0.9 DU which is well within the retrieval uncertainty showing that for most scenarios the spatio-temporal offsets with respect to ozonesondes are minimised.”

--Line 25: “ ... whole timespan of the operation of the European satellites...”
changed

--Line 26: “...since it is the only...”
changed

Page 5

--Line 2: “Possible reasons for the biases are...” The paper is filled with these statements. It would be good to know if this is the problem. Could you test your hypothesis by applying the same cloud algorithm to both retrievals? I realize that requires working with the instrument teams, but even a limited test application could prove useful. The differences that appear between the two panels of Figure 1 are striking. To me, this subject is more interesting than the one that is the current main focus of the paper.

It would be desirable to have the same cloud algorithm for all instruments. However, any bias from the cloud algorithm has been removed by the harmonisation process. We believe that the different spatial resolutions of the instruments is more important (GOME :320 x 40 km², SCIAMACHY: 60 x 30 km², and GOME-2: 80 x 40 km²).

--Figure 1. It occurs to me in looking at the upper panels that perhaps it would be worthwhile to separate the lower plots into “over land” and “over water” components. Visually, It would also be helpful to the reader if the plots were rotated 90 deg. so that the latitudes ran up and down the page as they do in the top panel.

In order to keep it simple we leave the figure as is. The land-sea contrast can be clearly seen in the 2D plots. The error bars in Fig. 1 (line graphs) mainly reflect the longitudinal variation possibly due to land-sea contrast. The line graphs have been changed so that the y-axis is the latitude.

--Line 8 – 9: “This behavior may be explained by the short time of common operation...” Another undemonstrated hypothesis. How could you test this assumption? And if it’s true, is not your transfer standard idea (i.e., reference to SCIAMACHY) compromised?

The larger variation in the bias with latitude in GOME data is most likely due to the short overlap period (10 months, from August 2002 to June 2003 (when GOME lost its global coverage)). For GOME-2 the overlap with SCIAMACHY was more than 5 years, making the latitude dependence smoother.

Page 6

--Line 3: "...ozonesonde data, it seems reasonable..."
changed

--Figure 2. What is important about this plot is how little of the area is actually statistically significant. Perhaps you should reverse and only mark with "x" those cells that ARE statistically significant.

Also, remind the reader, what fraction of the cells are statistically significant? This result is key in your argument that you can use a constant offset to "correct" the GOME 2 data, but you do not seem to make much of a point of that in the text.

Figure 2 has been changed, showing with x the statistically significant grid boxes. It is clear from the figure that the vast majority of the grid points are statistically insignificant and there is no need to specify.

Line 14, page 9 on the following modified sentence: "Scenario 6 can also be rejected due to the fact that the drift in GOME-2 correction offset at 81% of the grid-boxes is statistically non significant. "

We have added a line plot to Fig. 2 to show that the drift is not significant. The reviewer is correct that this is the main reason that a drift correction is not needed as already mentioned in the text.

--Line 12: "...but only a subset of them." Which subset? Which years?

.....of the ITCZ (no cloudy data available in the western Pacific)

The sentence has been removed since the explanation is given the sentence before.

--Sect. 2.3: This seems to be an important part of the paper, but frankly, I do not find it well motivated. Why 6 scenarios? Have you exhausted all possibilities. Does the reader need to know the details or simply your recommendation for the best approach to harmonize the data – with a discussion of the other approaches you tried and how they inform your estimate of the component of the calculated trend uncertainties that arise from the harmonization process itself?

We think that this is one of the most important results of this paper. We show here that the hamonisation procedure (merging) is one of the largest error sources of the trends. In the six scenarios we checked different reasonable assumptions on how to handle the differences between the individual instruments. The trends derived from the various merged dataset show larger differences than the statistical uncertainty from the trend regression applied to one of them. This is usually neglected in other studies.

To make this section a bit shorter as also suggested by Reviewer #2, we moved

this figure showing the maximum trend difference among all six merging scenarios to the supplementary material. Fig. S2 shows that the mean differences in trends from all pairs of merged datasets is about 2 DU/decade, exceeding in most cases the uncertainty from the single data regression.

Page 7

--Table 1. Not sure what to do with this table. Why are the stations in the order in which they appear? What is the table communicating? Here's where my earlier comment about SHADOZ reprocessing becomes relevant: are you using the reprocessed sonde data? How would these results change if you did? You are integrating the sondes to 200 hPa. How does that compare to the altitude used for the CCD approach?

Table 1 showed comparisons between integrated ozone columns up to 200hPa from 9 tropical ozonesonde stations from the SHADOZ network (version V05) with 6 different possible merging scenarios of Tropical tropospheric ozone columns from GOME, SCIAMACHY, and GOME-2. The tropospheric ozone columns retrieved with our CCD algorithm are adjusted to 200 hPa using climatological values (Leventidou et al., 2016).

The change in the differences of tropospheric ozone columns (up to 200 hPa) between collocated CCD_results and SHADOZ for the stations of Paramaribo, Am. Samoa, Hilo, and Fiji due to changes from SHADOZ V05 to V05.1R .

<i>Station</i>	<i>Scenario 1</i>	<i>Scenario 2</i>	<i>Scenario 3</i>	<i>Scenario 4</i>	<i>Scenario 5</i>	<i>Scenario 6</i>
<i>Am. Samoa</i>	<i>0.6</i>	<i>1.0</i>	<i>0.6</i>	<i>0.0</i>	<i>0.7</i>	<i>4.3</i>
<i>Paramaribo</i>	<i>1.8</i>	<i>0.7</i>	<i>1.7</i>	<i>2.7</i>	<i>1.8</i>	<i>-1.1</i>

Following the comments of the reviewer the comparison has been updated including the newest version of SHADOZ data (for two stations). The stations of Fiji and Hilo have been removed from the paper as discussed earlier. The following text has been added: "Fiji (18.1S, 178.4E) station is not included in the comparison because it is highly influenced by air coming in from mid-latitudes and the upper troposphere (Thompson et al., 2017). Hilo (19.4N, 155.4W) is influenced by volcanic out-gassing with high SO2 emissions, resulting in negligible ozone concentrations at the boundary layer. Therefore, this station is also not included."

The order of the stations has changed to alphabetical, and the title and the table has changed as follows:

"Mean differences (in DU) between merged TCO data, retrieved with the CCD method using six possible harmonisation scenarios, with integrated ozone columns up to 200 hPa from nine SHADOZ stations. The stations marked with asterisk present data from the newest reprocessed (V05.1_R) version (Thompson et al., 2007; Witte et al., 2017). The regions where the merged scenarios have the smallest biases with the ozonesondes are marked with bold. Scenario 1 has the smallest mean bias for all the stations."

<i>Station</i>	<i>Scenario 1</i>	<i>Scenario 2</i>	<i>Scenario 3</i>	<i>Scenario 4</i>	<i>Scenario 5</i>	<i>Scenario 6</i>
----------------	-------------------	-------------------	-------------------	-------------------	-------------------	-------------------

<i>Am. Samoa*</i>	<i>-0.89</i>	<i>-0.92</i>	<i>-1.99</i>	<i>-0.61</i>	<i>-0.93</i>	<i>4.59</i>
<i>Ascension</i>	<i>0.03</i>	<i>-0.14</i>	<i>-0.77</i>	<i>-0.42</i>	<i>-0.6</i>	<i>0.03</i>
<i>Java</i>	<i>-0.11</i>	<i>-0.12</i>	<i>-1.12</i>	<i>-0.54</i>	<i>-0.55</i>	<i>-0.11</i>
<i>Kuala Lumpur</i>	<i>-1.81</i>	<i>-2.12</i>	<i>-2.12</i>	<i>-2.14</i>	<i>-2.48</i>	<i>-1.78</i>
<i>Nairobi</i>	<i>1.81</i>	<i>1.10</i>	<i>1.80</i>	<i>1.48</i>	<i>0.74</i>	<i>1.84</i>
<i>Natal</i>	<i>0.56</i>	<i>0.63</i>	<i>-0.21</i>	<i>0.22</i>	<i>0.28</i>	<i>0.57</i>
<i>Paramaribo*</i>	<i>-2.98</i>	<i>-2.95</i>	<i>-3.02</i>	<i>-4.11</i>	<i>-4.34</i>	<i>-0.11</i>
<i>Mean bias for all stations</i>	<i>-0.48</i>	<i>-0.64</i>	<i>-1.06</i>	<i>0.87</i>	<i>-1.13</i>	<i>0.72</i>

The results with the updated ozonesonde data do not differ significantly from our earlier results. Nevertheless, we present our results now using the updated V05_R1 ozonesonde data for the available stations.

Page 8

--Line 7: "The same occurs for the..." delete the comma.
changed

--Line 9: "...the scenarios that can be confidently rejected according to this comparison are..." I'm not sure I have confidence that any scenario can be rejected until I know more about the sonde data you used and the altitudes used for the satellite tropospheric column amount.

The updated ozonesonde data to not change the conclusions. The tropospheric ozone columns from the sonde data were calculated exactly as the satellite data up to 200 hPa.

The text has been changed as follows(page 9, line 14):

" Although the comparison between the TCO from the individual harmonised scenarios and the ozonesonde data does not favor clearly any harmonisation scenario, the scenarios that can be confidently rejected are scenarios 3, 4 and 5 where GOME data are corrected with respect to SCIAMACHY since the overlap period between GOME and SCIAMACHY is very short (10 months, 8/2002-6/2003). Scenario 6 can also be rejected due to the fact that the drift in GOME-2 correction offset at 81% of the grid-boxes is statistically non significant. Lack of significant drifts in the comparison between GOME-2 and SCIAMACHY over the overlapping period shows that the data records are quite stable. Finally, scenario 1 (no drift corrections and bias correction for GOME-2) has the smallest mean bias with the ozone sondes (-0.4 DU). For these reasons, scenario 1 has been selected to be the preferred harmonisation scenario for merging the TCO datasets."

Page 9

--Line 1: "where α is the offset ..."
Changed

--Line 12: "...to persist into the next month."

changed

--Line 16: define "AR(1)"

The sentence has been changed as: "Therefore, the first order autocorrelation of the noise (AR[1]) is included in the model, as explained by Weatherhead et al. (1998)."

--Lines 25ff: "Nevertheless, all scenarios shown in Fig. 3 agree that there is a positive trend..." My quick read of Figure 3 is that very little of the map shows statistically significant trends. As to the fact that one appears to exist "over the southern tropical Atlantic Ocean" and a couple of other sites cited by the author, I am not sure what to make of it. The authors provide no explanation for such trends or why they might exist. I would find this more compelling if the authors could simplify the presentation, show the best correction scheme, show the best estimation of uncertainty (including that resulting from the harmonization scheme), and then spent some time in the text discussion what the resulting trend data showed and why. In its current form, I find the presentation more confusing than compelling.

Figure 3 e) shows the regions where the statistically significant trends calculated using the preferred merging scenario exceed the maximum difference of the trends among all six merging scenarios (now Fig S2) and can be reported with the highest confidence.

Page 11

--Line 8: "... following very well a Gaussian distribution." You don't show that in the paper. But I'm not sure what to make of it, either. Are you saying that there is no signal anywhere on the map? What do you mean by "the noise is random?" What is the noise?

This sentence has been removed as it is out of context here.

--Line 10ff: "This result is in agreement with Ziemke et al. (2005) and Ebojie et al. (2016)..." What periods did they examine? What data did they use? Are there any implications from the fact that it does not appear to have changed from their analyses to your analysis? What new have we learned from your analysis?

The refereed sentence (now page 11, line 30) has been modified as follows:

"The mean tropospheric ozone trend is in agreement with Ziemke et al. (2005) (using solar backscatter ultraviolet (SBUV) and Total Ozone Mapping Spectrometer (TOMS) version data from 1979 to 2003) and Ebojie et al. (2016) (using SCIAMACHY limb-nadir-matching (LNM) observations during the period 2003–2011) who also indicated insignificant and near zero global trends in the tropics, although their analysis was based on different datasets and covered shorter time periods. "

Page 12

--Line 10: "Figure 5 summarizes the tropical tropospheric ozone trends ..."

changed

--Line 25: "...may still be an artifact of the data-set." Is there a way to know?

The sentence (now page 11, line 18) has changed as follows: " The negative trends appearing in a region at the northern latitudes (Caribbean sea and northern Pacific) may be an artifact of the data-set (low sampling of data, 54 out of 240 months of data)."

--Line 31ff: Just to be clear, you selected your regions based on where you found statistical significance? That led to larger regions that then had statistically significant trends?

Yes, we selected the regions in order to have large number of grid points with significant trends for highlighting.

--Your Table 2 shows some impressive trend results. You then follow that with a list of possible factors that led to the trends (anthropogenic NO_x, population, energy consumption, biomass burning, changes in meteorology, dynamical oscillations, stratospheric intrusions) and you cite some prior works that have made these suggestions, but you provide no evidence within this paper for the proximal cause (or causes) in each of the regions you list in Table 2, nor is there really any justification for the selection of the boundaries of those regions other than they produce significant trends. If, as the title of this section suggests, mega-cities are responsible, it seems the regions might have been more narrowly defined. It would have been nicer to select regions based on a hypothesis and then identify the existence of significant trends (or not) to accept or reject that hypothesis.

The conversation about the possible reasons for the noticed trends has been moved in the conclusions. This section now summarises the areas where we observe statistically significant TTCO trends and compares these results with other studies.

Page 16

--Line 10: "Despite the fact that might appear to be..." What might appear?

The sentence has changed as follows (page 14, line14): "However, the observed trends over the northern and southern tropical latitudes (18° –20° in SH and NH) should be generally interpreted with caution because they are influenced by low sampling of data due to the movement of the ITCZ, which reduces the cloudy data during local winters and makes the above cloud ozone column (ACCO) retrieval difficult, violating in some cases the invariance of the ACCO per latitude band."

--Lines 22ff: "...cloudiness and humidity which contribute to photochemical O₃ loss..." I think the missing factor identified in the Morris et al., 2010 paper was significant lightning production, which they hypothesized led to NO_x production and O₃ loss in the absence of sunlight. The presentation here is a bit oversimplified. Deep convection alone can loft relatively low O₃ concentrations from near the surface (especially over the sea) to the upper troposphere. Those decreases are not "loss" but reductions resulting from transport.

We removed most of the discussions on possible causes as we can only speculate on it. In the Summary (page 18, line 34) we briefly mention that we cannot attribute the observed changes in tropospheric ozone as numerous factors may contribute the trends (production, loss, transport). Only with the help of modelling data one can disentangle the various factors.

Page 17

--Line 2: “winter” – what does “winter” mean in the tropics? Perhaps it’s better to identify seasons by months rather than such ambiguous names.

Changed

--Line 3: “as NO₂ over North America and Europe may have affected the O₃ trends...” moved the comma.

Changed

--Line 8ff: “Possible reasons...” There’s a whole list of possibilities here with no conclusions or evidence to support any one (or combination) of them.

The speculations about possible reasons that are responsible for the observed trends has been removed from the text. Instead a paragraph has been added in the summary where the complexity of trends' interpretation is discussed.

--Line 10: “...water vapor in the troposphere accounts for one of the most important...”

Changed

--Lines 10 – 12: “An increase in vertical convective patterns over the tropical oceans may result in lower ozone mixing ratios in the upper troposphere...” True if lofting low ozone from the surface. If lightning is present in the convection, however, you might see enhancements.

Thus, the influence is unclear.

See our earlier reply above (has been removed)

Page 18

--Lines 23ff: I would replace all of this text with a table. No need to write it all out.

Section 3.3.3. and the discussion about trends in mega cities has been removed from the paper as also suggested by Reviewer #2.

Page 19

--Table 3. Like previous tables, what is the logic of the order of cities in this table? How do the periods of study for Heue, Ebojie, Schneider, Hillboll, and this work compare? What impacts do differences in study periods have on interpretation of the results? The data in this table appear to have been compiled using 2.5 X 5 deg boxes. That’s roughly an area 250 km X 500 km in the tropics. Can you actually see signatures from megacities spread out over such a large area? For a control, should you also compute trends around cities that have not

grown (perhaps ones that have shrunk) or that have reduced emissions just to see if they behave any differently than the ones you list here?

See previous reply

--Line 3: "The derived tropospheric trends clearly show that tropospheric ozone increase is not proportional to ..." If you're going to make this claim, I think you need to show the population data, perhaps in Table 3, and the proxy you're using for industrial activity as well.

See previous reply

--Lines 8 – 9: "The degree of tropospheric ozone change strongly depends on the NO₂ amount..." As the second half of this sentence correctly relates, it depends on the relative NO₂ and VOC concentrations. I think I would get rid of the word "strongly" in this sentence.

See previous reply

Page 20

--Line 21: "... since the uncertainties in the trends are larger..."

See previous reply

--Line 24: Cite the uncertainties associated with the trends published in Ebojje et al.

See previous reply

--Line 31 – 32: "They might be linked to..." You list a whole bunch of possibilities. Has anyone shown the specific relevant link for your study? If not, how can you test these hypothetical influences?

The speculations about possible reasons that are responsible for the observed trends has been removed from the text. Instead a paragraph has been added in the summary where the complexity of the interpretation of the trends is discussed. It is out of the scope of this paper to attribute the trends to specific processes.

Page 21

--Line 4: "...tropical latitudes (> 18°N and S)..." What range?

The sentence has changed as follows (page 18, line13): "The most important limitation in interpreting the observed trends over the northern and southern tropical latitudes (18°–20° in SH and NH) is the low data sampling at these latitudes. "

--Line 10: "It has been shown that tropospheric ozone increase is not linearly related..." I'm not sure this study has shown that result conclusively or persuasively.

This text has been removed (mega cities)

--Line 18: "...the fact that their retrieval reaches up to the tropopause ..." That seems like an important factor. How different are the retrievals? What impact do the differences have on your results/homogenization scheme?

It is true that we cannot directly compare the trend results but, as mentioned earlier, we use them as indications of the range of the estimated trends.

--Line 28: "...(expected lifetime of the Sentinel 5 precursor satellite)." No need to introduce an abbreviation in the last paragraph.

changed