Response to Dr. M. E. Koukouli

The authors would like to thank Dr. M. E. Koukouli for her comprehensive review and for helping us to improve our manuscript substantially. Below, please find our detailed response to each one of the reviewer's comments.

1. Line 9, Page 1: Maybe you can use : "'also in", because being a two places at the same time sort of defies many major laws of physics. :)

We agree with the reviewer that the use of two current addresses may be confusing. We have only one current address in the revised manuscript.

2. Line 15, Page 1: I am rather concerned about this precise phrasing and would prefer to alter it, it gives an erroneous first impression on how this "corrected" dataset is all about.

We agree with the reviewer. We have rephrased this in the revised manuscript: *"…The GOME and GOME-2 data are "corrected" relative to the SCIAMACHY to produce a self-consistent dataset that covers the period 4/1996-9/2017…".*

3. Line 18, 19 and 21, Page 1: I do not think that you can use e.g. here since the numerical findings you quote are the results of your work, right? I would delete the e.g. from this location and the others also highlighted.

We agree with the reviewer. We have removed "e.g." in the revised manuscript.

4. Line 32, Page 2: Maybe you can update with this: Munro, R., Lang, R., Klaes, D., Poli, G., Retscher, C., Lindstrot, R., Huckle, R., Lacan, A., Grzegorski, M., Holdak, A., Kokhanovsky, A., Livschitz, J., and Eisinger, M.: The GOME-2 instrument on the Metop series of satellites: instrument design, calibration, and level 1 data processing – an overview, Atmos. Meas. Tech., 9, 1279-1301, https://doi.org/10.5194/amt-9-1279-2016, 2016.

The paper of Munro et al. (2016) is now cited.

5. Line 5, Page 3: This is not such a general reference to use for describing the two sensors, you could you the Munro et al., 2016, reference or also:

Hassinen, S., Balis, D., Bauer, H., Begoin, M., Delcloo, A., Eleftheratos, K., Gimeno Garcia, S., Granville, J., Grossi, M., Hao, N., Hedelt, P., Hendrick, F., Hess, M., Heue, K.-P., Hovila, J., Jønch-Sørensen, H., Kalakoski, N., Kauppi, A., Kiemle, S., Kins, L., Koukouli, M. E., Kujanpää, J., Lambert, J.-C., Lang, R., Lerot, C., Loyola, D., Pedergnana, M., Pinardi, G., Romahn, F., van Roozendael, M., Lutz, R., De Smedt, I., Stammes, P., Steinbrecht, W., Tamminen, J., Theys, N., Tilstra, L. G., Tuinder, O. N. E., Valks, P., Zerefos, C., Zimmer, W., and Zyrichidou, I.: Overview of the O3M SAF GOME-2 operational atmospheric composition and UV radiation data products and data availability, Atmos. Meas. Tech., 9, 383-407, https://doi.org/10.5194/amt-9-383-2016, 2016.

We thank the reviewer for her suggestion. We replace the Wang et al. (2017) paper with Munro et al. (2016) which is a more general reference.

6. Line 17, Page 3: The wording "first time" refer to the usage of the four specific sensors or the technique? I would be a bit weary in using such a phrase after spending a whole paragraph above stating other relevant efforts.

We agree with the reviewer, the phrase "for the first time" has been removed in the revised manuscript.

7. Line 19, Page 3: Updated also,

Levelt, P. F., Joiner, J., Tamminen, J., Veefkind, J. P., Bhartia, P. K., Stein Zweers, D. C., Duncan, B. N., Streets, D. G., Eskes, H., van der A, R., McLinden, C., Fioletov, V., Carn, S., de Laat, J., DeLand, M., Marchenko, S., McPeters, R., Ziemke, J., Fu, D., Liu, X., Pickering, K., Apituley, A., González Abad, G., Arola, A., Boersma, F., Chan Miller, C., Chance, K., de Graaf, M., Hakkarainen, J., Hassinen, S., Ialongo, I., Kleipool, Q., Krotkov, N., Li, C., Lamsal, L., Newman, P., Nowlan, C., Suleiman, R., Tilstra, L. G., Torres, O., Wang, H., and Wargan, K.: The Ozone Monitoring Instrument: overview of 14 years in space, Atmos. Chem. Phys., 18, 5699-5745, https://doi.org/10.5194/acp-18-5699-2018, 2018.

We thank the reviewer for noting that, the Levelt et al (2006) reference has been replaced from Levelt et al. (2018).

8. Line 8, Page 4: I assume that there exists a main paper that describes the algorithm? please add the reference here.

We cite the Boersma et al. (2004) paper here in the revised manuscript.

9. Line 23, Page 4: Is there a report/paper that details this methodology? for e.g. is there a minimum of days that are required for the equivalent month to be calculated? is there an effective day consideration? also, where all SZs there a report/paper that details this methodology? for e.g. is there a minimum of days that are required for the equivalent month to be calculated? is there an effective day consideration? also, where all SZAs used? how about locations with high albedo? was the associated error included in the averaging? how about negative trop NO2 columns which are also reported in the nominal data? and so on. As used? how about locations with high albedo? was the associated error included in the averaging? how about negative trop NO2 columns which are also reported in the nominal data? and so on.

We thank the reviewer for giving us the opportunity to clarify this. In the end of Sect. 2.1 we added the following phrase in the revised manuscript describing the method that was followed and the various filters that were applied when averaging.

"…When averaging, each observation is weighted by its fractional area (%) within the grid cell. For each valid observation, the cloud radiance fraction has to be less than 50% (cloud fraction less than about 20%) and the surface albedo not higher than 0.3, while observations with a solar zenith angle higher than 80^o are filtered-out. In addition, there is no limitation in the number of observations used, negative columns are taken into account, and the observational error is ignored in the averaging process (e.g. Schneider et al.. 2015).…"

10. Line 24, Page 4: Have these datasets been used previously in similar studies, or validation studies, or trend studies, or intercomparison campaigns, etc? if so, a brief mention is worth here.
This is the standard averaging method has been applied in a number of studies in the past. Here we cite Schneider et al. (2015) who used the same averaging method for the same version (TM4NO2A v.2.3) of data.

11. Line 25, Page 4: I have a major concern regarding the methodology provided below, which I am sure stems from the fact that the authors have not wished to increase their article in length by explaining in detail but I consider it paramount: since they already work on monthly mean gridded data, i.e. 0.25x0.25 deg, how do they justify in post-correcting for the different pixel sizes of the original instantaneous measurements? isn't this correction already taken care of by the original gridding where "When averaging, the observations were weighted by the size of the overlapping surface defined by the pixel and the corresponding grid cell." [verbatim from page 4, line 22.]

I am guessing that the first correction, CF1, has more to do in harmonising the datasets not in spatial resolution [horizontal resolution as the authors state] but in general differences due to observation geometry, spectrometric/radiometric differences, FWHM differences, etc.

I expect this topic to be discussed clearly and explained adequately in the next version of the manuscript.

We thank the reviewer for giving us the opportunity to clarify this. Indeed, when gridding the data, the original swath measurements are weighted according to the size of the overlapping surface defined by the pixel and the corresponding grid cell (see details in the answers above). However, this does not impact the fact that the information included in a larger swath corresponds to a larger area. Hence, the gridded data which are produced from larger pixels (e.g. 320 x 40 km^2 for GOME) will be of "lower resolution" than the ones produced from smaller pixels (e.g. 60 x 30 km^2 for SCIAMACHY) and the resulting maps will be much smoother, despite the fact that the grid cell size is the same (0.25^o x 0.25^o). As the GOME nominal resolution is nearly 3 times lower than that of SCIAMACHY at the horizontal dimension, we may assume that each grid cell of the GOME gridded data despite having the same size with SCIAMACHY grid cells, will correspond to an area nearly 3 times larger. The latitudinal dimension is considered to be close in both the datasets (40 km vs 30 km) and also close to the latitudinal dimension of the gridded data (0.25^o). This is why we used a boxcar algorithm with an averaging window of 13 x 0.25^o (3.25^o) in the longitudinal direction only, similarly to Geddes et al. (2016) to smooth the SCIAMACHY data (CF1). Our results are very close to the ones seen in Hilboll et al. (2013) where a more detailed pixel by pixel approach was followed. Taking into account this and the discussion above we prefer to preserve the term "spatial resolution correction" in the revised version of the manuscript. A paragraph explaining the whole reasoning has been added in the revised manuscript.

12. Line 1, Page 5: Also at this location, I suggest you alter this phrase. No one can really reproduce what an instrument would have seen where it in orbit, and functioning, longer.

We agree with the reviewer. We have removed this phrase in the revised manuscript.

13. Line 4, Page 5: What does this VCDsc stand for?

We mention in the revised manuscript: *"…the SCIAMACHY monthly gridded VCD data (VCD_{SC}) were smoothed…"*

14. Line 7, Page 5: It would be beneficial to explain why this correction is performed in the horizontal resolution only, since the satellite pixels are not entirely aligned in the east-west direction but, depending on the sensor, have a different geospatial direction.

We understand fully the reviewer's point. However, the fact that satellite pixels are not entirely aligned in the east-west direction having a different geospatial direction is ignored following Hilboll

et al. (2013) and Geddes et al. (2016), considering this of minor importance compared to the correction that has to be applied in the horizontal direction.

The following phrase is added in the revised manuscript: *The correction is applied in the horizontal dimension only as the along track dimensions are close in the two datasets (40 km vs 30 km) and also close to the latitudinal dimension of the gridded dataset (0.25^o).*

Please also look into our answer No.11 above and the corresponding paragraph we added in the revised manuscript.

15. Line 9, Page 5: I am assuming that this factor is calculated on molec/cm2. Which means that, for small changes in absolute numbers between VCDsc and VCDsc-sm, this may result in a rather un-physically large number. How do you deal with this issue? any limits imposed?

So, you are basically saying that you spread the SCIAMACHY observations to a pseudo-GOME pixel size and then you de-seasonalise them and then you multiply the CF1 factor to the GOME data in order to create a pseudo-GOME product on the SCIAMACHY pixel size?

We thank the reviewer for giving us the opportunity to improve our manuscript here. CF1 is dimensionless as it is the ratio of VCD_{SC} and VCD_{SCsm} on a climatological monthly basis. As discussed in the revised manuscript to avoid having unreasonably large CF1 values due to very low tropospheric NO_2 levels, CF1 was set equal to 1, in cases of VCD^{SCsm} lower than 0.1 x 10^{15} molecules cm^{-2} which corresponds to SCIAMACHY's precision.

Indeed, this is the general idea.

Line 14, Page 5: I understand why you would have to show these maps in the supplement, instead of the main text, however you definitely have to discuss them, explain what is seen, what was found, was there a seasonal pattern [which is not shown in these annual averages, for e.g. that summertime was systematically over or wintertime was over.] Since you are applying this technique on a species that has a very clear seasonal variability, such discussion is paramount. Also, in the legend of Fig S1, you note:

"A value of 1 was used in cases where the mean
tropospheric NO2 VCD was lower than SCIAMACHY's precision (0.1 x 1015 molecules cm-2)."

This is information that should appear in the text, alongside a comprehensive discussion on what a less than unity value signifies. I suggest that you can add one of these four figures in the main text.

Also, I was surprised to see patterns in the CF1 that appear to alter in the latitudinal direction and not the longitudinal direction, which is the one "corrected" for. How do you discuss this? please add in the text.

We added a new figure (Fig. 1) where we show the changes that the original GOME data undergo from step to step and discuss the correction factor patterns trying to connect them with areas of high and low tropospheric NO_2. This is a large improvement in our paper. In addition, the CF1 patterns have been plotted on a global and regional basis on a monthly and annual basis. Now the reader may get an idea about the seasonal variability of the CF patterns which are pretty persistent throughout the year as discussed in the revised manuscript. The CF1 patterns are given in high resolution in the supplement along with similar figures for CF2 and CF3. The paragraph has been enhanced so that the readers can find various important details about our method.

For the limit value of 1, please see our answer above.

As discussed in the revised manuscript CF1 exhibits characteristic spatial patterns with values greater than and lower than 1 over and adjacent to pollution hotspots, respectively. The CF1 patterns are pretty persistent throughout the year.

16. Line 14, Page 5: Do you apply the CF1 to GOME-2 as well? I am guessing not. Then maybe you should not mention GOME-2 here as it appears slightly confusing.

We agree. We refer to the GOME period only in the revised manuscript.

17. Line 19, Page 5: Shift correction due to which SCIA/GOME difference? more detail is needed here to explain why this CF2 is applied and what exactly it is.

The details are given in a couple of lines below that point *"…The shift correction factor (CF2: 1 value for each grid cell) is equal to the difference between the two datasets for the common period and was calculated on a grid cell basis (Eq. 4) similarly to Geddes et al. (2016)…"*. However, we added a few lines connecting this correction with the work of Hilboll et al. (2013). Following, Hilboll et al. (2016) who used a trend model that explicitly accounts for a level shift between the two instruments and for a change in the amplitude of the seasonal variation, we applied a shift correction (step 2) and a seasonal amplitude correction (step 3) successively, on top of the spatial resolution correction (step 1). The goal here is to account for instrumental biases that were not corrected with the spatial resolution correction. The method follows the reasoning of Geddes et al. (2016) who applied a shift correction but for annual data.

18. Line 20, Page 5: Which two datasets? SCIA VCDsc and GOME VCDcg1? SCIA VCDscsm and GOME VCDcg1? ... and GOME VCDg?

These two paragraphs have proven quite difficult to follow, since the reader has to go back and forth between this part of the paper, Appendix A and the supplement. I strongly suggest to the authors to move the equations here, and at least one of the Figs S1 to S4 and Figs S5 to S8 in this part of the text so that the reader can follow seemlessly the methodology and benefit from its discussion.

We added VCD_{SC} after SCIAMACHY: *"…were compared against SCIAMACHY data (VCDsc)…"* and the corresponding equations from the Appendix were placed at the end of the paragraph.
The equations that correspond to each correction step are now at the end of the corresponding paragraphs within the text and not in an Appendix.

19. Line 25, Page 5: Normalized to what?

We rephrase this paragraph also including the equations in the text.

20. Line 27, Page 5: Normalized to what?

The same as above.

21. Line 33, Page 5: As you know, GOME2A has been suffering from degradation effects and shows different levels for its atmospheric retrievals than GOME2B. Also the GOME2A pixel size has changed during the time period chosen. How are these issues dealt with/compensated for?

The GOME-2A and GOME-2B data were assumed to be of equal quality and resolution within the calculations. As 80 x 40 km^2 (GOME-2A before July 2013 and GOME-2B) is equally different to the SCIAMACHY resolution like 40 x 40 km^2 (GOME-2A after July 2013) the authors decided not to apply any corrections and the GOME-2A and GOME-2B data were averaged on a monthly basis. We mention that in the revised version of the manuscript.

22. Line 6, Page 6: Due to the systematic lack of....

Corrected.

23. Line 13, Page 6: What constitutes a major trend reversal as far as NO2 is concerned? this is a very important detail which have to be discussed and the choice explained.

We acknowledge that this phrase might be puzzling for the reader and hence we decided to omit it. The method reports only one reversal (where S(t) minimizes). There is not much possibility that there were more than one significant reversals (secondary S(t) minima) within the period studied here. However, in the future when there will be three or four decades of data there may be a need for the detection of more than one reversals. This is why in the last sentence of the paper we write: *"…the need to develop similar methods in the future that will be able to incorporate both morning and afternoon measurements (e.g. from OMI and TROPOMI)* ***and detect more than one trend reversal points*** *in improved tropospheric NO_2 products (e.g. QA4ECV v.1.1, Zara et al., 2018 and references therein) is acknowledged…"*

Also, the trend reversal method is now described in more details and it is placed within the text and not in an Appendix.

24. Line 16, Page 6: A general comment on the discussion of the methodology: even though the authors mention that they chose to have the equations and such like in the appendix so that the text quickly enters the results section, I find it very hard to follow [and be convinced by] the methodology since it is a constant back and forth to the appendix and the supplement. I strongly suggest that they re-think this strategy, that they add the equations in the main text, and also include an example of how the GOME [for e.g.] VCD alters between the nominal and the corrected levels for the three corrected VCDs. The annual CFs may not be so important, especially since the authors simply give the plot without discussion, but for the reader [and potential user of this new dataset] it is important to see how the original satellite data alter.

We thank the reviewer for giving us the opportunity to do this change in the revised manuscript. As discussed above, the detailed description of the method has been placed in the Methodology Section within the text. We agree that this will make it easier for the reader to follow the various changes that the data undergo within each step of the method. In addition, a figure (Fig. 1) has been added with the tropospheric NO_2 patterns for the whole GOME period from the original (VCD_G) data (a), from the corrected in step 1 (VCD_{GC1}) data (b), from the corrected in step 2 (VCD_{GC2}) data (c) and from the corrected in step 3 (VCD_{GC3}) data (d).

25. Line 19, Page 6: The multi year average tropospheric...

Corrected.

26. Line 19, Page 6: .... -GOME-2 dataset ... [space missing]

Corrected.

27. Line 21, Page 6: I am afraid that the choice of colour bar does not permit such a careful look, or it is a problem of providing the figure with sufficient analysis. I suggest you change the colour bar to the more typical, for e.g., http://www.temis.nl/airpollution/no2col/no2month_tropomi.php?Region=9&Year=2018&Month=02 which allows for the shipping tracks to show clearly.

We agree with the reviewer and hence we have changed our colobar. Now, the ship tracks can be seen more clearly. Our original figures are in high resolution and we will make sure in close collaboration with the copyediting department of the Copernicus that they will appear perfectly in the final version of the paper.

28. Line 25, Page 6: Maybe you could also add other works such as: https://www.nature.com/articles/srep35912; Krotkov, N. A. et al. Aura OMI observations of regional SO2 and NO2 pollution changes from 2005 to 2015. Atmos. Chem. Phys. 16, 4605–4629 (2016); https://www.tandfonline.com/doi/full/10.1080/01431161.2018.1430402; https://agupubs.onlinelibrary.wiley.com/doi/full/10.1002/2017GL076788; https://www.atmos-chem phys.net/17/9261/2017/

We also cite the works of de Foy et al. (2016) and Krotkov et al. (2016) which refer directly to the three urban clusters.

29. Line 28, Page 6: I also suggest that you add references to other works that have observed these hotspots from space, a short review of Scopus results would provide you with numerous choices.

We have added some references here including global maps with tropospheric NO_2 also suggesting that one should also look into the references given there.

30. Line 1, Page 7: maybe you can write "various socioeconomic changes", if you do not wish to mention specifically what types of changes.

We added "socioeconomic"

31. Line 5, Page 7: If the reason you are showing both these figures is to compare to the work of Schneider et al. then you definitely need to add a nice long comparison between your work and theirs. IF you do not plan to add such a comparison, then I would exclude Fig2a and add zoomed-in plots showing Europe/Africa/East Asia/etc, i.e. so that one can see also visually the findings you have in your tables.

Furthermore, since you have already discussed the trend reversal, how do you justify showing this Figure? for e.g. at locations where you found a trend reversal which trend do you depict in this plot?

We thank the reviewer for giving us the opportunity to clarify this. First of all, our goal is not to compare our results with Schneider et al. (2015); however, it is important that our results are similar to that of Schneider et al. (2015) despite the fact that we use a much longer period in our analysis and not a single sensor like they did. This is indicative of the fact that the sources are persistent during the period we study and despite the trend reversal that some regions experienced, the sign of the trends for the whole satellite period has not yet changed (the significance has changed in some cases).

As this is the first global map presented using more than two decades of tropospheric NO_2 from satellites we prefer to keep this figure as is. We have produced regional plots with the trends but we believe that it is better for the reader to have the whole picture.

When preparing the manuscript we also thought of keeping only the second panel (Fig. 3b) with the statistically significant results. However, we decided that we should include the non-statistically significant trends which they still show a kind of tendency. In addition, we have gone through comparisons of the trends with data from the recently released QA4-ECV product (http://www.qa4ecv.eu). We have seen that despite the fact that the trends are largely similar in the two datasets, the QA4-ECV, trends are much smoother over the oceans and in some cases the tendencies are opposite (see Deliverable 6.3: http://www.qa4ecv.eu/node/9). As this new dataset is going to be used in the future for trend analysis studies, we believe it is very important to include a map with the all the trends regardless of their significance. Unfortunately, keeping the Fig. 3a and indicating statistical significance (like e.g. in Pozzer et al., 2015) with a symbol (e.g. + or ·) is not possible due to the high resolution of the data.

When writing the paper at first place we decided to show the full period trends regardless of the fact that there may be a reversal. The calculation of the trends for a 2-decades period that we focus here period is not wrong, it gives us an indication of the general tendency of the tropospheric NO_2 levels during the whole period of measurements. However, the detection of trend reversals and the calculation of the trend for the period before and after is just more accurate. We strongly believe that the presentation of the full-period trends and then the trend reversals is the optimal way of stating that our results are similar to previous studies with shorter periods (and maybe one single sensor), but now it is time to start taking into account the trend reversals. With our approach one may also see that despite the fact that a trend appears to be insignificant there are two different periods (before and after the reversal) with significant trends of opposite sign.

We will make sure that the figure appears large enough and in high resolution in the final manuscript so that each detail can be seen.

32. Line 8, Page 7: Shouldn't the precision of the GOME and GOME2 sensors also worry you? what levels are those at? maybe 0.1x10^15 is a bit too optimistic?

We decided to keep a uniform precision of 0.1 x 10^{15} molecules cm^{-2} as the SCIAMACHY data are used as the GOME and GOME-2 data are "corrected" relative to the SCIAMACHY data. The same reasoning was followed by Hilboll et al. (2013).

33. Line 8, Page 7: It would also be interesting if you were to actually compare you work with that of Hilboll et al., 2013, since you actually used their technique.

We agree with the reviewer that a detailed comparison would be very interesting; however, this is not within the scope of the current paper. Here, we compare our findings with Hilboll et al. (2013) qualitatively rather than quantitatively and our results are indeed close to theirs. However, taking into account that we did not apply the same method but our method was based on their reasoning there may be differences.

34. Line 9, Page 7: I am not sure I follow your mathematical knowledge here, why does the cut-off value of 0.1x10^15 mol/cm2 mean 2 decimal places for the trend result and not 1 decimal place, for e.g.?

The 2 decimal places refer to the -0.0037 x 10^{15} molecules cm^{-2} yr^{-1} value that is well above the precision value (0.1 x 10^{15} molecules cm^{-2}). So, any trend above this "background" trend of our self-consistent dataset may be considered "real" and the use of two decimal places definitely ensures that.

35. Line 12, Page 7: Add here the actual numbers, I assume that they appear on one of the Tables further below, but it useful to have them here. Also, how do your trends compare to other trend studies? in numerics.

We have added the required information in the text.

"…America (the region of Mexico city). ***The trend values over these areas are higher than 0.05 x 10^{15} molecules cm^{-2} yr^{-1}, with a maximum value of 2.18 x 10^{15} molecules cm^{-2} yr^{-1} appearing within the BTH urban cluster in eastern China.*** *On the contrary, strong statistically significant negative trends appear over the largest part of the U.S. (especially the eastern U.S. and the state of California), western and central Europe, Japan and Taiwan in south-eastern Asia and the region around the Johannesburg-Pretoria conurbation in South Africa.* ***The absolute trend values over these areas are higher than 0.05 x 10^{15} molecules cm^{-2} yr^{-1}, with a maximum value of 1.40 x 10^{15} molecules cm^{-2} yr^{-1} appearing close to Los Angeles city in the eastern U.S****…"*

36. Line 20, Page 7: You definitely need numerics here, I am sorry to have to say. How similar is similar?

We refer here to the patterns and not the trend values per se. We acknowledge that the use of word "similar" here is not proper and hence we rephrased this sentence. Comparing the trend values with values from other studies quantitatively is not as trivial as it may seem because they refer to different periods. We therefore did not proceed to a quantitative comparison here.

"…In general, the trend patterns here resemble the ones appearing in previous satellite-based studies for shorter periods (e.g. van der A et al., 2008; Schneider et al., 2012; Hilboll et al., 2013; Krotkov et al., 2016)…"

37. Line 24, Page 7: Is this per annum or per decade? please specify.

This is the percentage decrease or increase (relative to the fitted mean for the first year) for the whole period. We give this information in the revised manuscript wherever these values appear.

38. Line 27, Page 7: Please compare these findings to other major works in literature, either satellite works or ground-based/in situ findings.

The % changes given here are not directly comparable to that of other works. This has to do with the fact that the trends are calculated relative to a different period. E.g. our % trends are calculated relative to the fitted mean for the first year (see van der A et al., 2006) while Schneider et al. (2015) used the whole period of SCIAMACHY as a base period and Hilboll et al. (2013) used 1996 as a base year. Taking this into account and our answers to comment 33 and 36 we refrained from comparing quantitatively these results with results from other works.

39. Line 29, Page 7: Indeed, which is now worrying that you are moving into stating that the trend already shown is not one trend but two. How do you explain this?

As discussed above we decided to show the full period trends regardless of the fact that there may be a reversal. The calculation of the trends for a 2-decades period that we focus here period is not wrong, it gives us an indication of the general tendency of the tropospheric NO_2 levels during the whole period of measurements. However, the detection of trend reversals and the calculation of the trend for the period before and after is just more accurate. We strongly believe that the presentation of the full-period trends and then the trend reversals is the optimal way of stating that our results are

similar to previous studies with shorter periods (and maybe one single sensor), but now it is time to start taking into account the trend reversals. With our approach one may also see that despite the fact that a trend appears to be insignificant there are two different periods (before and after the reversal) with significant trends of opposite sign.

40. Line 4, Page 8: I strongly suggest you break down these figures into sub-figures that show zoomed-in, as was shown for e.g. in Schneider and van der A, 2012. That way you can also discuss them more easily.

We thank the reviewer for giving us the opportunity to improve our figures. Zoomed-in global figures are provided in the revised manuscript. We excluded areas that did not present any reversals and at the same time we preserved all the interesting information into two single panels. As this is a highlight figure of the paper we will make sure in close collaboration with the copyediting department of the Copernicus that it will appear large and clear enough in the final version of the paper. All the figures are plotted in high resolution and we will make sure they appear perfectly.

41. Line 8, Page 8: This is precisely the point I cannot follow, how do you show the positive trend in Fig 2a and then in Fig 3a show the trend reversal?

Please see our answers to comment 39 and comment 31.

42. Line 17, Page 8: between

Corrected

43. Line 21, Page 8: What do these standards mean, in numbers, i.e. what are the new levels of emissions permitted?

The maximum allowed amount of on-road vehicle NOx emissions was reduced by 50%. This information is included in the revised manuscript.

44. Line 21, Page 8: This refers to NOx? VOC? particles?

This refers to a number of restrictions (e.g. a ban on older polluting cars) that were implemented in specific cities in China (e.g. Beijing) rather than to specific emissions. We have rephrased this in the revised version of the manuscript.

"…Stricter regulations were implemented on a city level for on-road vehicles (e.g. a ban on older polluting cars in Beijing)…"

45. Line 25, Page 8: (see Fig 3a.)

This was changed.

46. Line 26, Page 8: Maybe the quotes are not necessary.

Quotes were removed.

47. Line 28, Page 8: ... China, large....

Corrected.

48. Line 28, Page 8: I think that you have to discuss a bit on India's air quality here, and not to assume that all readers are well-versed in the details of the air over there. Much like you did for China, even though India does not have a five-year plan. What are the sources of pollution over there? what is already known? does this add to you finding a trend reversal in all those locations? Hilboll et al.., 2017, cannot be the only source of reference in this point.

We thank the reviewer for giving us the opportunity to enhance this paragraph in the revised manuscript (see below). In addition, a number of additional papers are now cited.

"… Similarly to eastern China, large parts of India experienced a reversal from positive to negative trends mostly in 2011. On the contrary, areas in central-southern India experienced a reversal from negative to positive trends at some point in the period 2000-2006. India experienced a population growth of ~37% (relative to 1996) during the period 1996-2017, mostly in urban areas, which was accompanied by a gross domestic product (GDP) increase of ~29% (World Bank, 2019). NO_x emissions generally increased as a result of large-scale urbanization (rural population decreased from ~73% of the total population in 1996 to ~66% in 2017), industrialization and economic growth, energy production, industry and transportation being the main contributors to the emissions (Ghude et al., 2013 and references therein). The Indian economy started developing at much higher rates after 2002 (World Bank, 2019) which might explain the observed negative to positive trend reversals appearing in the years 2000-2006 over specific regions (e.g. increase of tropospheric NO_2 in the greater Ballari region due to the rapid growth of the steel industry, especially after 2006). India's economic growth experienced a slow-down after 2011 (GDP still increased but at a lower rate) which might explain part of the observed positive to negative trend reversals over specific areas. Hilboll et al. (2017) also observed a stagnation of tropospheric NO_2 over India, attributing it to a combination of a slow-down in Indian economic development, the implementation of cleaner technology (e.g. Bansal and Bandivadekar, 2013), meteorological factors (see Voulgarakis et al., 2010) and changes in tropospheric chemistry. However, it has to be noted that the way all these parameters may influence the tropospheric NO_2 levels and trends over India is pretty complicated and should be studied in more detail in the future…"

49. Line 31, Page 8: So you mean that somewhere in those six years some areas in C-SE India showed a trend reversal? is this significant? is this something one may use? How much did the population increase within the ~21 years you are studying? how about demografics? how many farmers turned into city people? and so on.

We have rephrased that (see answer to comment 48). We mean that there is a reversal at some point within the period 2000-2006. As discussed several times in the manuscript we report a trend reversal only when the trend for the whole period before the reversal year or the trend for the whole period after the reversal year is statistically significant at the 95% confidence level. So, yes the result is statistically significant. Several numbers concerning the population statistics are given in the revised manuscript (see previous answer).

50. Line 32, Page 8: You mean that the steel industry became as stronger presence in that region? please re-phrase and also explain the steel industry emissions of NOx.

Yes, we have rephrased this sentence (see answer to comment 48). To manufacture steel requires very high temperatures for smelting and processing and the related combustion processes lead to emissions of NO_x.

51. Line 2, Page 9: You definitely need to expand on a few more details on this study, what does "in accordance" mean? numerics on the trend reversals from the in situ measurements are definitely a good idea to strengthen your case.

We have added a sentence here: *"…which is in accordance to NO_2 ground concentration measurements. More specifically, in Cuevas et al. (2014) a continuous drop of surface NO_2 is seen after 1999-2000 over the two cities…"*

52. Line 6, Page 9: Investigate further in either scientific references or policy making reports and add your sources to this statement.

We have added a sentence here and two related reports.

"…(EEA-APFS-Spain., 2014). Following the European Union directives, Spain introduced its First National Emission Reduction Program in 2003 setting stringent combustion emission standards (IEA, 2017). This was afterwards updated and revised leading to the Second National Emission Reduction Program in 2008."

53. Line 7, Page 9: Investigate further in either scientific references or policy making reports and add your sources to this statement.

The same as above.

54. Line 8, Page 9: Maybe there are other, more appropriate references to add here for the economic crisis of 2008 over Spain?

This is a highlight paper that discusses the effect of economic crisis on Spain NO_2 levels and it is the most appropriate for what we discuss here.

55. Line 9, Page 9: Any explanations on this fact?

We cite here a report (EEA-APFS-Portugal., 2014) where it is shown that the NOx emissions decline after 2005 in line with our findings. The trends are similar to that of Spain. So it is probably the compliance with the EU environmental directives that explain the 2004-2005 trend reversals while the reversals in late 2000s- early 2010s are probably related to the financial recession.

56. Line 11, Page 9: ... the whole of Syria...

Corrected.

57. Line 13, Page 9: I am sorry to repeat myself, but you are not doing yourself justice with this Figure. Again I urge you to break it down to sub-figs, and change the colour bar.

We have addressed that issue by providing zoomed-in global figures (see answer to comment 40). Prior choosing the trend reversal colorbar we did several tests using different colors and combinations and this one was found to be the most appropriate. Probably the conversion to pdf affected the colors. Our original figures are in high resolution and we will make sure they appear perfectly in the final article.

58. Line 15, Page 9: Why should this be?

The emissions in neighbouring countries (Syria, Iraq) have decreased due to warfare and hence transported pollution is expected to be less. We decided to remove the sentence about the political/financial involvement of Iran as it is difficult to explain it in a line. We have rephrased

accordingly: *"…while a decrease of transboundary transport of NO_2 from neighbouring countries due to the warfare cannot be ruled out…"*

59. Line 21, Page 9: Which year? which magnitude of trend reversal? why? and so on...

The trend reversals appear in the early 2000s and are statistically significant. The reasons are probably the same with the regions mentioned in the sentence above. We have rephrased accordingly.

"…Within the Middle East there are also sporadic areas (e.g. in Iran, in Iraq, areas around the Persian Gulf, and areas around the east coast of the Red Sea in Saudi Arabia and the Nile River in Egypt) with a trend reversal from negative to positive in the early 2000s (2000-2003) probably due to changes in power generation, industrial, transport and shipping emissions (Krotkov et al., 2016) (Fig. 4b)…"

60. Line 23, Page 9: Are you not going to discuss this finding further? I would assume that there exists a wealth of references and relevant studies for the US at least.

We prefer to refrain from reaching conclusions about the reversals in Mexico and Africa as there could be a number of different reasons (from environmental policies to land use changes, see e.g. Andela and van der Werf, 2014, doi: 10.1038/nclimate2313). A discussion about the U.S. is given in Sect. 3.4 mostly focusing on California and Los Angeles.

61. Line 2, Page 10: This Figure is very informative and shares a wealth of information extracted from your study, however:

1. Avoid the use of special text effects [for e.g. shadow] for the titles/legends/etc of the figures. In my humble opinion it makes reading the text very difficult, see for e.g. the trend information on the upper right plot. I suggest you remove all these effects and add the plots using the highest resolution possible. For the final version of the article, I of course highly recommend you provide ACP proof teams with *eps versions of your plots.
2. You have to explain the following at this point [it could have been done further up in the text of course]:
a. why did you choose to calculate trends on normal monthly mean time series and not the de-seasonalised monthly means?
b. how do you actually prove that the separation of trends results [right plots] is correct and not the full time series trend [left plots.]? you truly have to actually show this somehow mathematically. Or statistically. Or otherwise, but in actual numerics.

1. We have updated the figure as requested. Such changes have been applied on similar figures in the manuscript. As discussed above, probably the conversion to pdf affected the colors. Our original figures are in high resolution and we will make sure they appear perfectly in the final article.

2a. All the trends are calculated using the Fourier-based method described in detail in the manuscript. The method accounts for the seasonal variability using a Fourier-based seasonal component and hence there is no meaning in deseazonalizing the data before the application of the method as the method more or less "deseazonalizes" the data before calculating the linear trends.

2b. As discussed in previous comments the calculation of the trends for a 2-decades period that we focus here period is not wrong, it gives us an indication of the general tendency of the tropospheric NO_2 levels during the whole period of measurements. The detection of trend reversals and the

calculation of the trend for the period before and after is just more accurate. Hence, the use of a "correctness index" is not within the scope of the current study.

62. Line 28, Page 10: Again, it is not clear to me why both these figures are needed.

Our reasoning is discussed several times above. We want to show that despite there may be a statistically significant trend during the last 2 decades in the meantime there was a trend reversal. It is not wrong to calculate the trend for the whole period, it is just more accurate to break it into two trends and we believe these comparisons are indeed very informative. With the passing of the years and if the reversal continues the trend for the whole period may not be statistically significant any more. We believe our findings would be valuable for future trend studies including 3 or even more decades of satellite data.

63. Line 5, Page 11: This paragraph is rather difficult to read. I suggest a Table with in this information at this point. Table S1 is fine left in the Supplement.

We agree with the reviewer that this paragraph is very informative. In order to save space only countries that exhibit a trend reversal are shown in Table 2, while results for all the world countries are given in Table S1 of the Supplement. Some of these values appear in Table 2 and all of them in Table S1 so putting one more Table that will share values with Table 2 would be too much. As these might be very useful not only for scientists but also for journalists, media or other users we prefer to keep these numbers here.

64. Line 17, Page 11: You had a small phrase here which I wanted to highlight, but deleted by mistake. In any case, my recommendation is to delete it.

Corrected.

65. Line 5, Page 12: As above, please make all the letters/numbers/texts in this figure normal letters and not shadowed/text effects.

Corrected.

66. Line 12, Page 12: You had only one for Spain, so please enhance this.

As discussed above more citations are given in the revised manuscript for Spain.

67. Line 13, Page 12: By how much?

The decline was by 18% relative to 2008 levels. This is given in the revised manuscript.

68. Line 14, Page 12: Annual?

We refer to the annual mean levels. We have corrected accordingly.

69. Line 19, Page 12: Which economic recession is this? the pre-2000 one? is it referred to by a specific name, like the 2008 one is called, the economic crisis of 2008 [for e.g.?]

Details are given in Sect. 3.2. For Argentina we may refer to it as the 1998-2002 great depression, while for Brazil, there is not any specific name. We changes *"economic recession"* to *"economic recessions"* to make clear that we had two different crises.

70. Line 21, Page 12: I find is a bit hard to believe that the Brazilian Governement starting working towards the Rio Games from year 2000... :)

It should not be necessarily connected to the years around 2000. The preparations started gradually the years after 2009 and this could enhance the positive trend appearing after 2000.

71. Line 32, Page 12: Again, as discussed a number of times already previously, I fail to see the reasoning behind showing the trend for the total period since you are planning on showing trend reversals. You could split this map into two, and show trends for the locations without trend reversal for the entire period and then a two-panel plot for the other locations giving both and negative and positive trend and then year of reversal you find. Furthermore, I suggest you do not use an infinite colour bar here but one with say, 34 or 68 colours, so that the actual differences between locations can be seen with the naked eye.

Please see our answers above. The use of less colors did not change the figure drastically. Hence, we prefer to keep it as is in order to be consistent with the trend maps on a global and country basis.

72. Line 4, Page 13: Comment as per the similar paragraph above.

As discussed above we agree with the reviewer that this paragraph is very informative. However, we believe that such statistics are very useful not only for scientists but also for journalists, media or other users and hence we strongly believe they should be included in the text.

73. Line 14, Page 14: Comment as above.

Corrected.

74. Line 8, Page 16: Unless you can verify this with financial figures [e.g. GDP, plant emission increases, vehicle numbers increasing] , it does seem a bit extreme, a positive trend starting in 2000.

Details are given in the text. The GDP decreased after 1998 and started increasing in 2002.

75. Line 22, Page 16: Reference missing.

We have corrected this. Actually the reference is Zara et al. (2018).

76. Line 22, Page 16: Maybe better to say "strengthened"?

We decided to keep "acknowledged" here.