We thank the reviewer #3 for his/her helpful comments that have helped improve the manuscript. Below, please find our responses in boldface for each comment. We have updated the manuscript addressing these comments.

Overview

The paper by Choi et al. compares aircraft (from several campaigns) measurements from two different instruments and ground-based PANDORA measurements to satellite observed NO2 columns from OMI. The paper aims to interpret the differences between the datasets and explores different techniques how the satellite observations can be compared to aircraft measurements. The authors look into the effect of using a more accurate NO2 a priori profile and the effect of correcting for the large satellite pixel size. Overall, I found the paper was well-written and scientifically sound, it fits well in the scope of Atmospheric Measurement Techniques. I have some suggestions (listed below), after addressing these, I would recommend the paper for publication in AMT.

Specific Comments

1) Abstract and Introduction: It would be helpful to include the time period of the campaigns in the abstract and introduction, just roughly, e.g. "...these campaigns took place between 2011 and 2016."

We have specified the years of campaigns in the abstract as suggested.

2) p. 9, l. 254: Include what filters have been applied to the OMI data. Here, you mention the row anomaly, but what other filters? I think it would be good to mention it here, instead of l. 369 "quality-controlled, cloud-free". What do you mean specifically with "quality-controlled"?

We agree. We have added the following sentence in the revised manuscript, and removed similar statement in line 369 of the previous version:

We use OMI pixels with cloud radiance fraction less than 50 % and quality flags suggesting good data.

3) The correlation is discussed for each location (in Sects. 2.3.2 and 3.2); I think that there might not be enough measurements for many of these for the correlation to be meaningful. Some of the high correlation appears for locations where there are only a handful of measurements (e.g. Korea), with 1 or 2 high column amounts. I think it would be better to just discuss the correlation for all measurements from all locations rather than separating them. I think it's ok to talk about the differences for each location.

We have modified the statements discussing correlations of individual campaigns with those with overall correlation, as suggested. The sentence in Section 2.3.2 (Page 10, Line 304) now reads:

The correlation ranges from fair (r=0.41, N=21) for MD and TX to excellent (r>0.9, N=36) for CA and Korea with the overall correlation coefficient of 0.53.

And the sentence in Section 3.2 (Page 12, Line 367) now reads:

The overall correlation coefficient between Pandora and the airborne NCAR and TD-LIF measurements are 0.94 and 0.91, respectively, with higher correlation in CO, TX, and Korea and lower correlation in MD and CA.

4) Table 3 (and A2-A4): Include the correlation and difference for all locations (total), also include columns for the sample size (N)- I know it is listed in Table 2, but it would be helpful to have this information all together in these tables. I believe from this it will be clear that maybe not too much weight should be given to the correlation for individual sites.

We have revised the tables as suggested.

5) Where are the spirals relative to Pandora? Are the spirals flown over the Pandora locations? Or how far is the distance? It might be useful to show maps (could be in the appendix), one for each of the five sites with the location of the spirals, PANDORA site and maybe (if it's not too busy) the pixel outline from OMI observations used for the comparison. This will be helpful for the discussion and interpretation of the differ- ences between the measurements taken on the aircraft, from the satellite and from the ground

Thank you for pointing that out. The surface sites hosting in-situ NO_2 monitors and Pandora instruments are within 2-5 km from aircraft spirals during DISCOVER-AQ and 10-20 km from aircraft ascents/descents during the KORUS-AQ field campaigns. During DISCOVER-AQ, the spiral diameter was about 4km on average. We have included this information in Section 2.1 (Page 4, Line 87) in the revised manuscript as below:

The P-3B aircraft made spirals of ~4km diameter whereas the DC-8 ascents/descents covered 10-20 km. Consequently, the distance between the ground and aircraft locations was 0-5 km during the DISCOVER-AQ and 10-20 km during the KORUS-AQ campaign.

We have also revised Fig. 4 showing the location of a surface site as an example to provide the idea of relative locations of ground, aircraft, and OMI measurements.

6) Fig.5: Where is this? Please include lon/lat ticks, and the location name in the caption.

This is for the Maryland campaign. We have revised the figure and the figure caption as suggested.

7) Fig. 9: The error bar is cut off from Olympic park. I would suggest changing the figure slightly (it is necessary to read the caption to know which colorbar applies, it is not immediately obvious to the reader): Maybe instead of using two different y-axes, it might be better to include a scale break for the Olympic park columns, or a logarithmic scale.

We have revised Fig. 9 using a Y scale break as suggested.

8) Table 2: It says in the caption "Pandora in parenthesis", many do not have a number in parenthesis, is that if the spirals are the same number as Pandora, or are no Pandora measurements available? This is not clear. To make it easier for the reader include parenthesis with the number of PANDORA measurements everywhere.

We agree with the reviewer. The table is now modified for clarity.

9) Did you evaluate the effect of changing the time difference slightly to a stricter or more relaxed criteria to 1.5h (1h or 30min, or 2h)? Are some of the outliers related to a large time difference?

We tested the effect of different time windows on the results. Narrower time windows improve temporal matching, but decrease the number of samples. Using wider time windows could help reduce the impact of outliers and found to slightly decrease the correlations, but the change is marginal. Our selected time window was intended to maximize the number of samples while reducing effects from diurnal variation of NO$_2$ profiles. We have added the following statement in the revised manuscript (Page 4, Line 114):

> **This time window of ±1.5 hour is selected to maximize the number of samples while reducing effects from diurnal variation of NO$_2$.**

10) I guess you only considered the OMI pixel that overlapped with the aircraft data, or did you average the OMI observations? Was this specifically mentioned somewhere? It should be explained.

We select only the OMI pixel that overlap with individual aircraft profiles. Changes made in Section 3.3 (Page 13, Line 377)

Technical Corrections
11) p.1 l.15: "very well, but. . ."

Done.

12) p.2 l.18 "but mostly to OMI's areal (>312km2) averaging" change to "but mostly to OMI's large footprint (>312 km2)."

Done. Changed as suggested.

    13) p. 5 l.138: "figures below" change to "Figs. X-Y." (Which figures? Please specify.)

Changed as suggested.

    14) p. 9, l. 252: Space is missing ". Since"

Done, thanks.