We thank the reviewer #2 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.

General Comments

This paper assesses NO2 measurements from two field campaigns for their usefulness in interpreting satellite remote sensing observations. The paper focuses on what these field campaigns can tell us about how spatial resolution and a priori NO2 profiles affect interpretation of satellite observations. These are important and actively researched questions in the satellite retrieval community. The paper is well written, and the analysis is generally well supported. I have a couple questions listed below, but otherwise I recommend this paper for publication.

Specific comments

1) In section 2.1.2, the authors describe how aircraft profiles are linearly interpolated to the surface to meet the surface monitoring measurements. Figure 2 indicates that the merged aircraft profiles decrease towards the surface, while both models show increasing NO2 towards the surface. It seems that the decrease towards the surface in the merged profiles is due to the interpolation – is there concern that the merged profile decreases towards the surface (which is generally unexpected over polluted areas), or that it has such a different shape than the model? AMF calculations for the satellite observations are sensitive to the profile shape in the lower troposphere, so what impact does this interpolation have on AMF calculations?

Main idea of Figure 2 is to show the nature of variability in observed and simulated NO2 vertical profiles over the entire domain for each individual campaign. Due to mismatch in both spatial and temporal sampling between models and observations, this figure is merely for a qualitative understanding, and cannot be directly linked to results presented in the later sections. Our analysis focusing on comparison of different data sets and examining the AMF effect uses more restrictive collocation (spatial and temporal). To address this concern, we have added the following statements (Page 3, Line 124):

> **Figure 2 also shows the nature of variability in observed and simulated NO2 vertical profiles over the campaign domains. The observed differences between the model and observations arise primarily from mismatch in both spatial and temporal sampling. Use of more restrictive collocation (spatial and temporal) applied for comparing different data sets in Section 3.1 and examining the AMF effect in Section 2.3.2 would have resulted in different vertical distributions.**

Moreover, the linear interpolation between the surface and lowest aircraft altitude (~250 m) is done to partially account for missing information. The missing information certainly represents a significant source of errors, and the interpolation adopted here is our attempts to mitigate some of these errors. We have discussed these limitations in Section 3.4 (Page 14, Line 424-432)

2) I was surprised to see the results discussed in the paragraph beginning line 304, which showed that monthly mean profiles capture the local variability as well as the daily profiles. Using monthly means rather than daily profiles would simplify future retrievals, so I would be interested to hear whether the authors think that this result is particular to the observed locations and seasons or if it can be applied more generally. For example, the only winter observations are in California (which isn't really that wintery), so could monthly mean profiles still be useful in these and other cases?

We agree with the reviewer. AMFs calculated using local, daily vs campaign average monthly NO_2 profiles show a good agreement with correlation coefficient of 0.9 (Figure 3), suggesting that local monthly mean profile can be a reasonable choice as discussed in the manuscript. Use of monthly mean profiles also simplifies retrieval process as the reviewer pointed out. We have discussed this aspect in Section 2.3.2 (Page 11, Lines 315-323). However, previous works (e.g., Laughner et al, 2018) have shown that errors form using monthly average profiles are as large as 30%. This may suggest that daily profiles that do not capture plume can potentially lead to considerable errors. This is the topic of active research, and should be pursued as more campaign data become available.

3) In section 3.2, the authors derive tropospheric columns from Pandora measurements by subtracting the OMI stratospheric column from the Pandora total column. I wonder whether this approach may be partially responsible for biases between Pandora and the aircraft observations. As a space-based instrument, OMI is more sensitive to stratospheric NO2, while Pandora has a greater tropospheric sensitivity. How do the authors account for differences in vertically-resolved sensitivity between the instruments? Also, what is the possible effect of subpixel variability in stratospheric NO2 within the OMI pixel? There's nothing in the discussion that describes potential errors in the stratosphere-troposphere separation.

Thank you for pointing this out. Use of OMI stratospheric NO2 columns to derive tropospheric columns from Pandora could impact the comparison between Pandora and aircraft observations. But, this is likely a minor effect because (1) OMI stratospheric NO2 estimates are fairly accurate following improvements and assessment of slant column retrievals by the European and NASA teams (e.g., Marchenko et al., 2015; Zara et al., 2018; Boersma et al., 2018); (2) contribution of the stratosphere on total column NO2 is small in polluted areas, where these campaigns took place; (3) the stratospheric NO2 field is relatively smooth and is unlikely to vary significantly within few kilometers; and (4) variation in vertical sensitivity for both OMI and Pandora measurements are accounted for in the retrievals. We have included the following statement in the revised manuscript (Page 12, Line 364).

> **The use of OMI stratospheric NO_2 columns to derive tropospheric columns from Pandora could impact the comparison between Pandora and aircraft observations; this approach is unlikely to be a significant factor over the polluted DISCOVER-AQ and KORUS-AQ campaign domains.**

4) Lastly, there's a recent paper by Judd et al (https://doi.org/10.5194/amt-12-6091-2019) that may be of interest to the authors (I am not affiliated with this paper, but thought it was relevant).

Thanks for the suggestion. We have cited Judd et al in the revised version.