Reviewer 2:

The authors investigate the variability of NO2 within hypothetical satellite footprint sizes based on high spatial resolution airborne imaging datasets. For this purpose, two different methods (random pixel sampling and spatial structure functions) are applied that provide consistent results. In addition, the authors also address temporal variability. The topic is of importance for the scientific community and fits well into the scope of AMT. However, I miss some information on the airborne dataset that should be included before publication.

Response: Thank you for your time and effort in reviewing our manuscript. We have addressed the comments accordingly. Please see below for details.

General comments

1. There is no explanation on the L2 retrieval of the NO2 tropospheric VCD. I think that you need to briefly describe, or at least reference to the retrieval settings. The auxiliary data used in the L2 retrieval also impacts on the observed variability (e.g. NO2 vertical profile shape, surface reflectance)

- What input data was used for the airborne measurements?
- Did you use a consistent retrieval for all regions (SMA, Busan, LA)

Response: We added the following description to Section 2.1 of the manuscript:

"NO_2 is retrieved from GeoTASO spectra using the Differential Optical Absorption Spectroscopy (DOAS) technique. The retrieval methods and Level 2 data processing are described in Lamsal et al. (2017) and Souri et al. (2020) for KORUS-AQ and in Judd et al. (2019) for SARP. Although beyond the scope of this work, it is important to recognize that assumptions made in the retrieval process (e.g., assumed vertical distribution of the NO_2 profile) could affect the final variability of the retrieved NO_2 fields."

2. I think that the study on temporal variability would benefit from consideration of the wind conditions (speed, direction, variability)

Response: Thank you for pointing this out. We analyzed wind fields from MERRA-2 reanalysis dataset for the 12 rasters over Seoul Metropolitan Area and 5 rasters in the LA Basin. We found that the wind conditions could be used to partially explain the patterns of the temporal variability, even though the impact of wind conditions on spatial variability and spatial SGV is small. We have included wind patterns of the MERRA-2 model surface level and 3 km (a.s.l.) in the supplement, and added discussion in the manuscript. Specifically, we have included the following statement to the manuscript:

"In addition, changing wind conditions (speed and direction; Figure S9) during the day can also lead to a shift in pollution pattern, and result in different pollution conditions for the same pixel at different time of a day. For example, Raster 1 of the 0609AM (9.17 local time) and Raster 2 of 0609PM (17 local time) are used to calculate TeMD for Dt equals 8 hours. The differences in wind conditions (Figure S9) and the pollution patterns (Figure 2) are large.".

Detailed comments & technical corrections:

3. Page 0 Line 0: What is the meaning of the bold polygon in Fig 1, Fig 2, Fig S1?

Response: The bold polygon represents the political boundary of Seoul. We added this information in the captions of Fig 1, Fig 2, and Fig S1.

4. Page 2 Line 65: also mention that comparison to in-situ observations is also difficult due to imperfect knowledge of the vertical profile

Response: We added the statement in the manuscript:

"… SGV introduces large uncertainties on top of the existing difficulty due to imperfect knowledge of the vertical profiles."

5. Page 2 Line 75: there are several more airborne instruments, which provide similar datasets. So these measurements are not really unique, see also P4.L126.

Response: We removed the adjective "unique" for GeoTASO in the manuscript.

6. Page 4 Line 130: Remove square brackets from citation

Response: We changed all the square brackets to parentheses for citation.

7. Page 5 Line 165: You state that you sample 10k hypothetical pixels. However, considering the area covered by the flights and pixel sizes of up to 25km2 it is not clear to me how many distinct samples are actually evaluated.

Response: Because we discarded a sampled satellite pixel if it is not covered by GeoTASO data for at least 75% of its area, the actual distinct sample sizes (~10% of all the samples) are much smaller than 10,000. For each hourly bin, there are ~1000 samples, therefore for a flight (typically 4 hours), there are 3200 samples for one satellite pixel size. The samples are sufficient as our sensitivity test indicates that the results do not change by halving the sample size. This information is included in the manuscript (Section 2.4).

8. Page 5 Line 167: Does -> do

Response: This was corrected in the previous revision process (i.e., minor revision before posting as preprint for interactive discussion).

9. Page 5 Line 169: How do you treat overlappings swaths from adjacent flight tracks? Do you

also account for temporal differences between these overpasses?

Response: As the GeoTASO data located closely in space may be sampled at slightly different times for the same flight, we separate GeoTASO data into hourly bins for each flight before pixel sampling in order to reduce the impact of temporal variability of the GeoTASO data within a single satellite pixel sample.

In addition, to quantify possible impacts of temporal differences in aggregating/grouping data samples used for our analyses of spatial SGV, we did two sensitivity tests that include different levels of temporal differences in the satellite pixel random sampling for spatial variability process.

(1) We tested the sensitivity of the results over SMA to sampling GeoTASO data with hypothetical satellite pixels grouped by complete flight, rather than grouping the data by time in hourly bins. The resulting patterns and relationships are similar, except that the normalized satellite SGV increases ~5% for pixels of small sizes due to the inclusion of temporal variability (Figure S7a).
(2) We then also tested the results for sampling satellite pixels by raster instead of within hourly bins. The results are again similar to Figure 4, except that the normalized satellite SGV increases ~1% for pixels of small sizes due to the inclusion of temporal variability (Figure S7b).

10. Page 6 Line 222: The SSF ais defined here follows Follette-Cook et al. (2015)

Response: We changed "is" to "as".

11. Page 6 Line 234: "SMA in the Discussion section"… Please include a proper cross-reference

Response: We changed "the Discussion section" to "Section 4".

12. S Fig S3: The labels are way too small. Please increase the font size or reduce white space between subplots.

Response: We reduced white space between subplots and increased the font size for Figures S3-S7.

13. Page 7/S Line 250/Fig S3: The differences between median and mean values seem to be much larger for the SMA region that for Busan and LA. Please discuss possible reasons and the impact on the normalized SGV.

Response: The mean values are larger than median values over SMA, while over the other two regions, mean and median values are relatively close. This is likely due to the high pollution level and extreme pollution events over SMA. Overall, we do not expect this to have a significant impact on the normalized satellite SGV. Because the high pollution level and extreme pollution events over SMA also lead to higher standard deviation (SD) besides higher mean. Higher SD and higher mean cancel out in the calculation of normalized SGV (the standard deviation of the GeoTASO data within the sampled satellite pixel divided by the mean of the GeoTASO data within the sampled satellite pixel; $SD_{pixel}/MEAN_{pixel}$). This is also consistent with our results– the pattern of

normalized SGV over the three regions are similar, even though they have different levels of pollution (Figure 4).

14. Page 7 Line 257: “discussed below”. Please include a proper cross-reference.

Response: We deleted the statement “that is discussed below”.

15. S6 Fig S6 Add a legend to the figure.

Mention in the caption that "red" corresponds to low values, and "blue" to high values. Consider using different colors, because “red” is also the color for the median.

Response: We added the legend.

When there are multiple boxplots in one panel (i.e., Figures S6 and S7), the color for the median is the same as the color of the box instead of red. For example in Figure S6, median of morning data is blue while the median of afternoon data is red). Therefore, there is no need to change to a different color.

16. Page 7 Line 265: What is “this relationship”

Response: We deleted the paragraph/sentence.

17. S S7 Add a legend to the figure

Response: We deleted the figure.

18. Page 8 Line 290ff: I am not sure if the threshold of ~10km spatial resolution can be generalized. It may be true for the regions investigated here. However, the spatial distribution of the NO2 field is also stongly affected by the meteorological conditions (strong winds lead to confined plumes, calm winds to high pollution levels above the sources) as well as the spatial distribution of the sources.

Response:

We analyzed wind fields from MERRA-2 reanalysis dataset for the 12 rasters (new Figure S9), and found that the wind conditions (both wind speed and wind direction) vary strongly among the 12 rasters. Related to the wind conditions, the spatial distributions of NO_2 field and pollution levels above the sources also vary strongly among the 12 rasters. We agree with the reviewer that the wind conditions can affect the spatial distribution of the NO_2 field and pollution levels above the sources. However, as we note in the manuscript (Section 3.1), the results show that the normalized SGV is not affected by pollution levels, and therefore less likely to be affected by wind field.

Nevertheless, we deleted the sentence to avoid potential confusion.

19. Page 9 Line 341: Why does a thicker PBL lead to stronger horizontal dispersion?

Response: During the daytime, increasing surface temperature leads to stronger vertical mixing and hence greater PBL height. In general, the greater vertical mixing is associated with stronger horizontal divergence at the top of the convective cell within PBL and hence potentially a stronger horizontal dispersion due to the divergence. We have revised the statement in the manuscript to reflect the explanation:

"As the day progresses, the PBL height increases (~1800 m during 15:00-17:00; Figure S9) allowing for greater horizontal mixing to take place."

to

"As the day progresses, the PBL height increases (~1800 m during 15:00-17:00; Figure S9) due to enhanced convection, which further induces a stronger horizontal divergence at the top of the convective cell and hence allows for greater horizontal dispersion to take place along with the divergence.".

20. Page 9 Line 348: What about changing wind directions? A change in wind direction would also lead to a shifted spatial pollution pattern, which consequently leads to a change in pollution levels over time above a certain location.

Response: Please see the response to Reviewer 2, Comment 2.

21. Page 10 Line 373: In Fig 6 you describe the increase of the mean differences of NO2 VCD with increasing time for the SMA region. The data over LA (Fig S11) does not show a similar behavior. Instead there is almost no change between Dt=4h and Dt=8h. Please provide possible explanations.

Response: The data over the LA is limited. Besides the limited data, different wind conditions over SMA and LA could be a possible reason for the difference in TeMD. We added the following discussion in the manuscript:

"For the LA Basin GeoTASO data, sampled hypothetical satellite pixels show TeMD increases at higher spatial resolution for the available Dt equal to 4 and 8 hours (Figure S11). However, TeMD is fairly constant at these two time differences which is different to what was observed over SMA (Figure 6). We note that with only 2 flight days of flight data, the GeoTASO data over LA is also limited, which may be the main driver of the difference. Besides the limited data, one possible reason is the different wind fields over the two regions. As mentioned previously, Raster 1 of the 0609AM and Raster 2 of 0609PM are used to calculate TeMD for Dt equals 8 hours over SMA. The differences in wind direction (Figure S9) for the two rasters are large (almost opposite in some cases). However, over LA, the differences in wind direction (Figure S12) for the two rasters (Rasters 1 and 3 for 0627 flight) are relatively small, compared to the differences over SMA.".

22. Page 10 Line 349: SSP? Do you mean SSF?

Response: The typo was corrected in the previous revision process (i.e., minor revision before posting as preprint for interactive discussion).

23. Page 10 Line 406: Are wind speeds of ~5m/s also representative for the measurement conditions of this study?

Response: We analyzed wind fields from MERRA-2 reanalysis dataset for the 12 rasters (new Figure S9). The averaged wind speed over the SMA domain region (upper left box in Figure 1) for the 12 rasters vary from ~1 m/s to ~3 m/s at the model surface level, from ~1 m/s to ~10 m/s at 3 km, and from ~2 m/s to ~17 m/s at 5 km. ~5m/s is within the range of the conditions represented by the 12 rasters.

24. Page 11 Line 425ff: What would be the dimensions of such a lookup table? Would you also consider the size of the city, the distribution of sources….?

Response: The three cities we studied have different levels of pollution and urbanization, city sizes, and PBL conditions. Their normalized SGV have a similar a pattern, indicating the pattern of normalized SGV may be generalizable to NO_2 VC over regions with different levels of urbanization and air pollution, and different PBL conditions. Therefore, in our future study and development of the lookup table by including more campaign data, we do not expect to include these dimensions. However, if the future study suggests such lookup table is not generalizable, we will alternatively provide the statistics for normalized SGV as a function of potential driving factors such as levels of pollution, city sizes, meteorological and seasonal conditions.

25. Page 14 Line 546ff: The reference styles are inconsistent.

* Some refereces have a DOI, others do not

* Some references have DOI as as clickable (blue) link, others do not

* Some refernces use doi:xxxxx, most others use https://doi.org/xxxx

Response: We unified the reference format according to the EGU's guide for reference format.