Reviewer 3:
We would like to thank you for your comments as we appreciate the time dedicated for this review and have made changes to the manuscript to reflect the suggestions. Individual comments from the review are bolded below with our responses in italics.

This paper by Judd et al. compares satellite-based TROPOMI tropospheric NO2 measurements with airborne- and ground-based Pandora measurements in the New-York City/Long Island Sound region. It contributes to the Sentinel-5P TROPOMI validation and is the first validation paper for the new satellite instrument with airborne campaign measurements which have a more spatially representativity than ground-based measurements. In addition, long-terme ground-based Pandora measurements are used and compared to the airborne and satellite based NO2 measurements. The strength of both reference measurements are used to evaluate TROPOMI tropospheric NO2 column densities. The evaluation found a low bias of the TROPOMI tropospheric vertical column (TrVC) compared to Pandora and aircraft tropospheric vertical column, more pronounced for aircraft than Pandora measurements. Although using a higher resolution a priori vertical profile for the TROPOMI data improves the low bias, there is still a low bias, especially for more polluted cases and further investigations are needed in future studies. Cloud retrieval effects are discussed. A new quality criterion was introduced which excludes pixel where the difference between retrieved cloud pressure and surface pressure exceeds 50 hPa to exclude pixels where cloud shielding occurred over cloud free scenes. These pixels compensate partially for the TROPOMI TrVC low bias but lower the correlations with reference measurements.
The paper is well written and of significance for the validation of the new satellite Sentinel-5P TROPOMI tropospheric NO2 measurements. Therefore, I recommend publication in AMT with minor revisions.

Specific comments:
Line 197: "All reference spectra were co-located with total column NO2 measurements from Pandora spectrometers: 5.6*10₁₅ molecules cm–2 at MadisonCT on June 30th, 5.7*10₁₅ molecules cm–2 at MadisonCT on July 2nd, and 6.2*10₁₅ molecules cm–2 at WestportCT on August 5th, with values estimated to be over 50% stratospheric."
What is done with the collocated Pandora measurements? How is the 50% stratospheric estimated?
The Pandora measurements are collected during co-located reference spectra scenes by the airborne spectrometer and are used to estimate the total column for our reference. The airborne spectrometer NO2 data has its own above aircraft (stratospheric) value estimated based on coincident TROPOMI stratospheric columns with the diurnal pattern from a climatology created with the PRATMO photochemical box model. That separation is needed in the calculation converting differential slant column to vertical column (e.g., the screenshot equation from Lamsal et al., 2017) with various other versions found in Judd et al, (2019) and Nowlan et al. (2018). I also added some text to reflect this calculation in Sect 2.3.

$$\Omega_v \downarrow = \frac{d\Omega_s - \Omega_v \uparrow \times A \uparrow + \left(\Omega_v{}^R \downarrow \times A^R \downarrow + \Omega_v{}^R \uparrow \times A^R \uparrow \right)}{A \downarrow}.$$

They are estimated as 50% by using the estimated airborne stratospheric column.

$$\Omega_{Pandora\ troposphere} = \Omega_{Pandora} - \Omega_{airborne\ stratosphere}$$
$$where\ \Omega_{Pandora\ troposphere} = \Omega_v^R \Downarrow \quad and\ \Omega_{airborne\ stratosphere} = \Omega_v^R \Uparrow$$

Line 292: What is the spatial coincidence criterion for Pandora comparisons to TROPOMI? Is it the nearest pixel, a mean, is the viewing direction considered?
It is the pixel in which the Pandora resides and viewing direction is not considered. This is clarified in Sect. 2.5.

Line 254: "All Pandora data are converted from total vertical columns to TrVCs by subtracting either the airborne or TROPOMI retrieved stratospheric columns for comparison purposes."

Is the Pandora converted with TROPOMI retrieved stratospheric column used for TROPOMI comparisons and Pandora converted with airborne for airborne comparisons?
*Yes and yes.*

How is the airborne stratospheric column retrieved?
The airborne stratospheric component is estimated using a stratospheric NO_2 climatology developed using the PRATMO (PRather ATmospheric MOdel) Photochemical Box Model (Prather, 1992; McLinden et al., 2000; Nowlan et al., 2016). The stratospheric column is bias corrected daily using TROPOMI NO_2 stratospheric vertical columns by calculating the average offset between the two datasets over the LISTOS domain for each day (ranging from $5x10^{13}$ to $6x10^{14}$ molecules cm^{-2}). To reflect this I reworded the sentence to say: 'All Pandora data are converted from total vertical columns to TrVCs by subtracting either the airborne <u>estimated</u> or TROPOMI retrieved stratospheric columns for comparison purposes.'

Line 450: Why was this feature only seen by this excluded Pandora?
It is by the nature of the plume, which is stated in Section 4, which we did observe with the airborne spectrometer. See below for an image of the large change in morning and afternoon and the structure of the plume i[...]
TROPO[...]
plume i[...]

Tropospheric NO_2 Column $x10^{15}$ molecules cm^{-2}
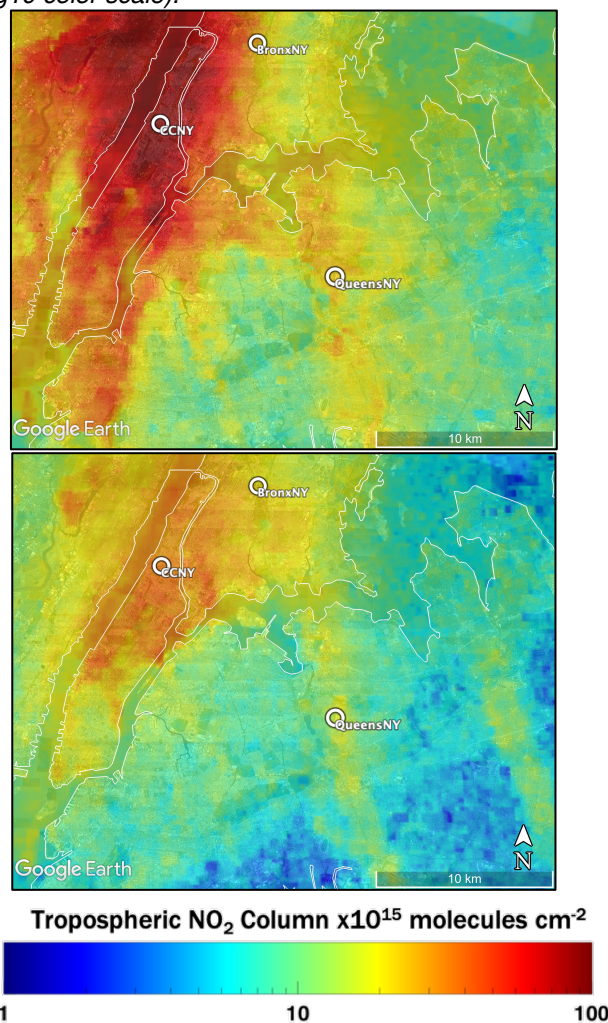
Airborne spectrometer NO2 TrVC during the morning (top) and afternoon (bottom) flights of July 2nd, 2018. The labeled circles indicate 3 Pandora locations, where CCNY is not included in this analysis (will be investigated for future work).

Line 571: Is there an explanation why the slope is much better and the correlation much worse when comparing TROPOMI and Pandora instead of TROPOMI and aircraft measurements?

The correlation is better for the airborne spectrometer because the TROPOMI sub-pixel variation is sampled by the airborne data. However, Pandora data is still subject to impacts from sub-pixel variability in this heterogeneous environment. It seems to be that the slope of the fit is also related to the sub-pixel heterogeneity. If you look at Table 6, you'll note that for sites where r^2 is greater than 0.5, the slopes range from 0.53 to 0.84.

Line 675: Lorente et al. did not used Pandora spectrometers for validation, they also found a low bias but with in-situ measurements.

You are right, the wording was wrongly reflecting Lorente et al. since they use in situ data. I broadened the scope of this paragraph now to include some other studies (including MAX-DOAS observational studies) as suggested by another reviewer.

Technical corrections:
Line 99: LISTOS is defined and used already in line 21 and 36.

I removed this definition of LISTOS. Now it should only be defined once in the abstract and once in the main manuscript.

Line 283: "TROPOMI NO$_2$ columns"
Better TrVC to be consistent to the other TrVCs in the sentence.

I see this point. However, during editing, this sentence had already changed in the next version and doesn't have this discrepancy anymore.

Line 372: "(Table S1, compare Row I to Row B) slightly improves the correlation (r$_2$ increases from 0.93 to 0.94)"
Row I is 0.94 and row B 0.92. Value 0.93 should be changed to 0.92 and order of "compare Row I to Row B" should be changed to "B to I" to make it consistent to the values order.

Thanks for catching the translation error in the r^2. This was fixed.

Line 420: "with large sub-pixel variation as indicated by the horizontal whiskers in the plot" There is a better explanation but some lines later (Line 433). This one could be replaced by the later one.
"The horizontal bars in Fig. 6 show the standard deviation of the subpixel airborne TrVCs within each TROPOMI pixel."

I removed the former phrase opting for the later discussion of spatial variation.

Line 556/Figure 9: Statistics are only listed in the table. It would be helpful for a better reading to have at least the statistics of the LISTOS time period data in the figure especially because these are much more discussed in the following than the statistics of the long-term TROPOMI-Pandora comparison.

I added the statistics for the LISTOS time period to Figure 9.

Line 651: r$_2$ of 0.89 should be 0.88 corresponding to the figure.

Updated the figure. R^2 is 0.885, which rounds up to 0.89. Thanks for catching this error in translation.

Line 714: r$_2$ of 0.84 should be 0.88

It should be 0.84 as I am referring to only the LISTOS timeframe for all Pandora sites (Figure 10a).

Line 722: and
corrected

Table 2: kg instead of lbs
Converted the weights to kg.

Table 3: Short explanation for shaded boxes
Added: 'Flights with shaded boxes are not considered in this analysis.' to the table description. Explanation is in the text as to why they aren't include: 'Only flights from 25 June – 6 September (13 flight days) are considered in this analysis due to availability of the high-resolution model data used to provide the a priori NO_2 profile shapes in the full vertical column retrieval (Table 1)'

Figure 1: Nine Pandora spectrometers instead of spectrometer.
Fixed

Figure 2: Explanation to horizontal and vertical bars with "variability at the time of measurement" is missing in figure caption.
For brevity it is described by the '($\pm$ min/max observation within a $\pm$ 5-minute window from the aircraft overpass)' and '($\pm 10^{th}$-90th percentile)'.

Figure 10: The period (LISTOS or extended long-term) of the used data is missing.
Good catch! Added this detail.

Figure 12: (a) Also for the LISTOS comparison only the extended stations are used
I clarified this figure caption now to ensure it states that it is only the four stations with extended temporal extent.

Supplement Line 53: "to remove the estimated of loss of sensitivity"
First "of" can be removed
This phrase was corrected.