Reviewer 1 Response:

We thank reviewer 1 for their time and thoughtful response to the manuscript. The responses to each comment are below where the authors' responses are bolded. New additions are in quotation marks and updated line numbers are included.

Changes made in marked Manuscript: Reviewer 1 Reviewer 2 BOTH

Reviewer 1:

This review is for egusphere-2025-5261, titled, Evaluation of Pandora HCHO and NO_2 with Airborne In Situ Observations. In this analysis, the authors evaluate Pandora HCHO and NO2 columns using in situ data from the NASA DC-8 during the AEROMMA period and also compare to TEMPO data. This evaluation is more a case study analysis rather than a statistical evaluation of the Pandora retrievals. Overall, the authors have done a dense analysis during these cases but is recommended for publication after a number of minor revisions are addressed as discussed below.

The analysis discussion going back and forth between NO2 results and HCHO results is confusing at times. Please work on addressing ways to make it clearer to the reader which product is being discussed, particularly in the California case (e.g., 3.1.2 is about NO2 but has a lot of discussion of HCHO).

Section 3.1.2 has been removed. New sections now read:

3.1 Case Study: Spiral over EdwardsCA Pandora

3.1.1 In situ Observations

3.1.2 Pandora HCHO and NO_2

The results do not get compared to the expected accuracy and precision of the Pandora and TEMPO. Consider adding this perspective to the analysis.

Based on discussions with PGN team members, formal accuracy and precision estimates for Pandora HCHO products are not yet well established. Existing studies suggest that uncertainties can be substantial and variable depending on atmospheric conditions and retrieval assumptions. This lack of well-constrained accuracy is a key motivation for this study, which aims to help characterize the performance and limitations of Pandora HCHO retrievals using independent airborne observations. A sentence has been added to line 103.

Line 17-18: This sentence only talks about validation of Pandora NO2 and ignores HCHO validation. Neither product from Pandora HCHO has been largely validated.

Sentenced has been rephrased to reflect this on lines 17-19

Line 26: The MAX-DOAS isn't sensitive to the viewing angle, but the comparison is due to the large spatial variance in NO2. The sentence is worded to imply that it's the technique but instead it's the characteristic about the variability and the mismatch spatially between measurements.

Rephrased to improve clarity on line 26.

Line 30-32: Are these stats relative to DC8 or TEMPO?

Compared to TEMPO. Now included in line 33.

Line 43-44: The MAX-DOAS also uses differential absorption. The sentences implies it does not.

Clarified in text line 46.

Line 45-46: The final sentence in this paragraph does not fit with the rest of the paragraph.

Removed

Line 58: describe what is meant by drift.

We have included a new sentence to clarify "MAX DOAS is able to better account for instrumentation wavelength drift by taking a new reference during every sky scan while DS retrievals rely on a fixed reference that may have been take up to a year apart from the measurement " (line 57-59)

Line 79: MMS needs a reference

Scott et al., 1990 has been included as a reference for MMS. (line 79)

Section 2.1: Please describe why more of the AEROMMA dataset was not used over Pandora sites (e.g., Whittier). It is likely criteria desired in the vertical completeness of the profile but

explain the criteria used for identification of these cases from the larger AEROMMA dataset. Are there any partial column comparisons that can be done at other sites?

Additional comparisons within the AEROMMA dataset are limited due to observational constraints. Whittier and other nearby Pandora monitors were either not in operation during the fight overpass time, were not performing long MAX DOAS scans, or the plane was too far away. There was a second spiral over EdwardsAFB in the second half of the AEROMMA mission, but the DC-8 did not go close enough to the surface for the profiling data to overlap with the Pandora max height. Profiling was done in Toronto, however the three Pandoras in that region were not close enough.

A sentence has been added in section 2.1 (line 84-86) to briefly discuss this.

Line 89: add "DOAS" before 'fitting'

Added (line 91)

Line 90: clarify that the sentence is about DS.

Clarified (line91)

Line 101: typo 'Blick'

Fixed (line 105)

Line 102-103: This data source is not the same as in the data availability section.

URL has been updated (line 107)

Section 2.3: This section mentions TEMPO NO2 but then does not explain why TEMPO NO2 is not used in this analysis like TEMPO HCHO. Please add why this is not included and justify why only HCHO data is considered.

This study focuses on HCHO because it has received significantly less validation in Pandora observations compared to NO_2, for which multiple validation studies already exist.

In addition, a complementary analysis of TEMPO NO_2 using in situ observations is currently being conducted separately. To maintain a clear and focused scope, we therefore limit this study to HCHO evaluation, where the need for validation is greater.

Line 225: clarify that the 6.6e15 value is from in situ integrated data.

Fixed (line 241)

Table 2: Clarify in the Table that DS NO2 has a climatological stratospheric value subtracted.

Table 2 description has been updated (line 234)

Line 237: clarify that it's MAX-DOAS if talking about the column but if talking about the profile, it doesn't appear to be 50% different throughout the entire profile.

The following sentences have been updated lines (249-252): "Pandora MAX DOAS NO_2 is 46% lower than the total in situ integrated column (1.2×10^{15} and 2.2×10^{15} molecules/cm^2, respectively). Excluding near-surface enhancements (red triangles in Fig. 3c) reduces the Pandora MAX DOAS NO_2 difference to 37% lower than the aircraft observations."

Line 249: add 'than in situ' after 'lower'

Done (line 263)

Line 251: clarify that it's 40% lower for the MAX-DOAS Pandora tropospheric column.

Done (line 265)

Line 257 and the preceding lines: there is also an uncertainty in the climatological stratosphere that would be an important factor, especially at such low column values.

The following sentence has been added to lines 272-274: "Additionally, there is uncertainty in climatological stratospheric NO_2 values, which is likely more pronounced at lower column amounts and may contribute to the observed discrepancies."

Line 261: the partial columns do not get anywhere near the tropopause, so this statement does not make sense.

Took out tropopause and used "higher altitudes" instead (line 278)

Section 3.2: Is there overlapping data in these flights? The flight summaries from the NOAA AEROMMA website talk about doing the pattern in reverse on the way back. If so, should these be separated? Otherwise, discuss why not or why only one pass is shown.

There is overlapping data in the flight maps. A sentence further describing the flights has been added (line 286-287) to reflect this: "During NYC1, NYC2, and NYC3 sampling began in New Jersey, continued through NYC and along the LIS into CT and then returned along the same route". The panels beneath the maps showing the time series of altitude colored by HCHO observations, along with the time stamps on the flight map itself helps to show the trend of HCHO.

Bottom of Page 12: State the maximum HCHO measured on NYC1 and NYC 2 to be comparative with NYC3 discussion. Winds are not mentioned either which is relevant to the discussion of Figure 5.

Max in situ HCHO for both days has been added to line 298. Surface wind discussion has been added on line 301. SFIG S7 also has been updated to include NYC1 and NYC2.

Line 309: Is this not also evident in the DS HCHO too?

Included DS in statement as well (line 331)

Paragraph starting on Line 321. As the AEROMMA aircraft had two passes of sites each day, is there any work that can be done to evaluate the 50% growth vs. 125% growth between DS and MAX-DOAS?

While the aircraft completed two passes each day, the total flight time in NYC only spans about 3 hours, all of which occurred in the afternoon. Since most of the change in Pandora columns began in the morning, we did not choose to include any diurnal component of the airborne data.

Line 329-331: While there is more data remaining, there would still need to be validation between cloudy and clear sky scenes for MAX-DOAS. Qualitative analysis is still valid but would need accuracy validation for quantitative analysis.

The following sentence has been added on line 363: "However, additional validation is necessary to assess the accuracy of MAX-DOAS retrievals under cloudy conditions, particularly given the potential impacts of cloud scattering and radiative transfer assumptions on retrieval sensitivity."

Figure 5: It is clear there is something wrong with Manhattan DS retrievals. Any indication what is going on there?

The following sentence has been added to line 349 to address this: "ManhattanNY-CCNY has been cited as a lower confidence Pandora in previous (Rawat et al., 2025), likely due in part to obstructed viewing from surrounding tall buildings."

Line 365-366: why are the buildings of NYC mentioned in this paragraph if no vertical profiles were collected over NYC? The sentence seems out of place.

Sentence has been removed.

Line 371: are in situ columns not extrapolated to the ground for the column calculation? They should be, but the explanation of the difference implied that it is not. If they are, then please better explain what is meant by the difference being from larger partial columns in the lower BL.

In situ columns are not extrapolated to the ground for the NYC comparisons. Instead, we only integrate both in situ and Pandora columns over a common range. We have added clarification on lines 395-397.

Line 371: Also the statement in this sentence isn't true for NYC2. Which is stated in the next paragraph, but it just makes the first sentence starting in Line 371 inaccurate.

Sentence has been fixed to say that only on average are the Pandora columns greater than the in situ (line 401).

Line 397-398: Why is the NewHavenCT profile on NYC3 binned differently and how?

NYC3 has only one Pandora *long* scan from NewHavenCT (other nearby Pandoras did not have valid data). There is no need to bin the scans for this reason. We have included more details in the figure 5 caption to explain this (line 429).

Line 400: what are the primary sources of HCHO here? Add references that state this is a factor in the NYC area?

The following lines (line 432-435) have been updated: In New York City in the summertime, the majority of HCHO is produced through VOC oxidation (Lin et al., 2012). Additional

HCHO directly emitted from combustion sources and land/water meteorological influences can result in a horizontal gradient of HCHO that must be considered when interpreting Pandora columns.

In the Pandora section, make it clear whether the tropospheric columns from the MAX-DOAS profiles are the same as the sum of the partial columns and if they're derived separately.

The following sentence as been added to line 97: "The sum of the partial columns beneath the maximum altitude is equal to the tropospheric column".

Figure 5 and Line 427: Are these TEMPO data that are collocated over the Pandora sites or are they over the whole domain shown in Figure 7? If the latter, this is not a valid comparison as the majority is over water where on NYC4 you show that all the high levels are over land. This is skewing the analysis. They should not be quantitively compared in this way.

The TEMPO columns in this analysis are averaged over the entire NYC domain in Fig. 2, rather than collocated directly with individual Pandora locations. We acknowledge that portions of this domain extend over the ocean, which may introduce a low bias in TEMPO HCHO columns due to generally lower marine values. These comparisons are therefore intended to provide qualitative context rather than strict quantitative validation, given the spatial mismatch between satellite pixels, Pandora observations, and aircraft sampling.

Due to the noise and uncertainty in TEMPO HCHO retrievals, a relatively large spatial domain is required for averaging; however, this choice introduces representativeness limitations that must be considered when interpreting differences. Two hourly HCHO retrievals from TEMPO are included at the bottom of this document to demonstrate the noise in HCHO columns.

A section has been added to lines 506-510 to address this.

The following sentence has been added to line 460 "These values correspond to column averages over the entire NYC domain as shown in Fig. 4.".

Fig. 5 figure caption has been updated (line 356).

Line 442: MAX-DOAS still needs validation under these partially cloudy scenes. But this discussion puts all the emphasis on the struggle of TEMPO or satellites in general with clouds. While there is data for evaluation of MAX-DOAS on this day from Figure 6, this appears to be after the clouds clear in the afternoon hours.

Thank you for this comment. Yes, we agree that MAX DOAS needs to be validated under cloudy conditions, even if the data passes the filtering assumptions. Lines 474-478 have been updated to say the following: "While TEMPO data lacks complete coverage under partially cloudy conditions, the MAX DOAS, which observes from the bottom-up, still provides valid data based on our filtering assumptions in the afternoon. However, additional validation checks are necessary to understand the accuracy of MAX DOAS under these cloudy conditions".

Figure 8: Please clarify in the discussion that each point represents a TEMPO scan time (if this is the case). However, it implies that TEMPO is averaged over the whole domain rather than at the Pandora sites. As stated above, this is not a valid method for quantitative comparisons. Data should only be considered over the Pandora sites and a spatial matching criteria should be stated.

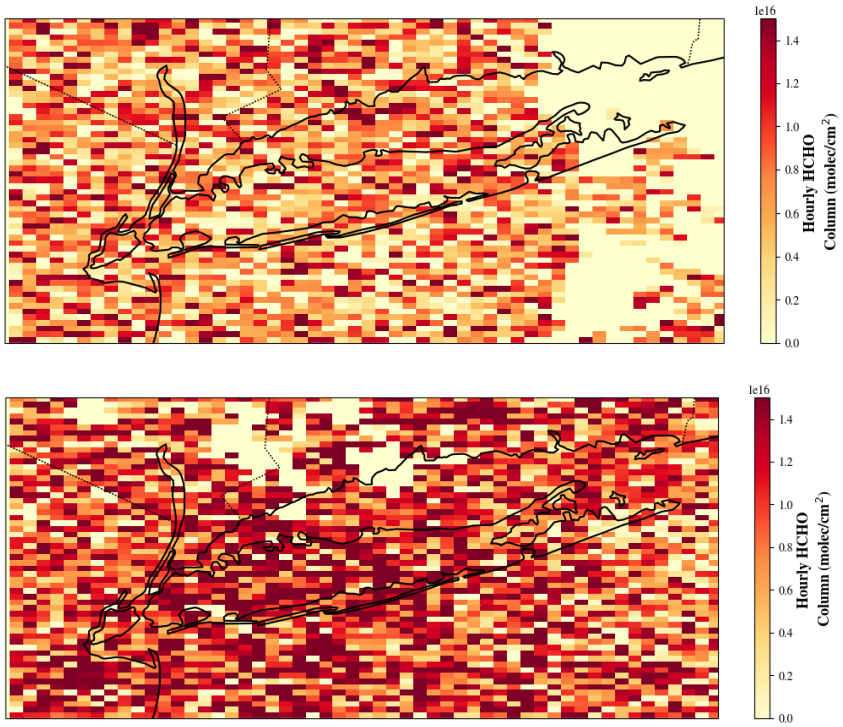
Fig. 8 figure caption has been updated. The following lines have been added (lines 506-510) "There is also potential bias associated with the choice of domain used for the TEMPO averaging. Due to the uncertainty and noise in TEMPO HCHO retrievals, a relatively large area is required for spatial averaging. However, portions of this domain extend over the ocean where HCHO columns may be lower than what is observed by the Pandoras. This may cause a low bias in the TEMPO HCHO."

Line 469 states a slope of 0.93 and the figure and the table do not even show this value. Same with the stated intercept.

Thank you for pointing this out. Values in the text have been changed to match the correct values in Table 3. (line 503)

Line 471: Why is there not a reason to believe the bias is present in Pandora and instead is stated as all on TEMPO? There is not enough evidence here to pinpoint which dataset is accurate, if any of them.

This is correct. We have revised the text to acknowledge that biases may exist in both the TEMPO and Pandora HCHO retrievals in line 506. Given the uncertainties associated with both datasets, this study does not attempt to attribute discrepancies to a single source, but instead highlights areas where further validation is needed.



TEMPO column HCHO in NYC at 8 am EDT (top) and 11 am EDT (bottom) on August 8th, 2023.