

Response to Reviewers

Evaluating vertically resolved Pandora formaldehyde retrievals using airborne and ground-based in situ measurements — Atmospheric Measurement Techniques

We thank both reviewers for their careful and constructive reading of the manuscript. Both reviews converge on a similar set of requests: more quantitative support for the comparisons shown, expanded discussion of prior literature, and a clearer accounting of how representative the four-site, two-campaign dataset is of the broader Pandora network. Because RC1 and RC2 raised many overlapping points, we summarize below the major revisions made in response. The point-by-point responses that follow refer back to the relevant summary item; comments unique to one reviewer are answered in full there.

Summary of Revisions

1. Scoping the operational recommendation: mechanism, evidence, and relation to Sebol et al. (2026)

The recommendation to adopt the revised height mapping rests on two independent grounds. First, a mechanistic explanation: the O_2 – O_2 -derived path length used by the operational mapping is best understood as an upper bound on how far a MAX-DOAS averaging kernel's sensitivity reaches, not the center of a near-symmetric, probability distribution kernel. Real MAX-DOAS sensitivity behaves like a saturating cumulative distribution function that rises from the surface, so treating the geometric proxy as a well-defined “midpoint” height builds in the low-altitude over-concentration this study identifies; a layer's own information content is the marginal difference between successive saturation heights, not a slice centered on a point. Because this is a property of the mapping's underlying assumptions rather than of any one site or season, the correction is expected to generalize beyond the sampled conditions. This argument is now articulated more clearly in Section 4.

Secondly: this dataset comprises, to our knowledge, the most direct evaluation of the PGN height mapping against independent ground truth to date, more so than the operational mapping's original (theoretically motivated, geometry-only) basis. We expanded the discussion of Sebol et al. (2026), whose AEROMMA-based comparison reports the same qualitative shape artifact but rests on a single true vertical spiral (Edwards AFB). Their four NYC-area flights had no true high-altitude spirals, sampled HCHO with standard deviations of order 1 ppb about comparable-magnitude means along level legs. As a result, the authors describe that comparison as “not ideal for evaluation of Pandora performance as differences may arise from air mass heterogeneity.” We link the shape artifact both studies report directly to this discussion and cite it as independent corroboration.

We also note that our recommendations are currently undergoing network-wide PGN internal evaluation. Additional considerations and areas for improvement may be uncovered via this evaluation—the next version of the operational retrieval will reflect the results.

2. Quantitative comparisons added throughout (Figures 4 and 8)

- DSCD fitting-window comparison ($N = 30,270$): the revised fitting window yields HCHO dSCDs consistently $\sim 2\%$ larger than the operational window; O_4 fit uncertainty is reduced by roughly 80% at every site-day, with the spread between revised and standard fits growing with airmass while remaining centered near zero.

Species	Revised vs. standard fitting window
HCHO	slope = 1.024, intercept ≈ 0 , $r = 0.995$
O_4	slope = 0.959, $r = 0.828$

- Figure 4 including regression (Theil–Sen) results: all spirals ($N = 31$), $y = 0.779x + 3.60 \times 10^{15}$; restricted to lowest sampled altitude ≤ 500 m ($N = 25$), $y = 0.820x + 2.90 \times 10^{15}$; excluding Essex, MD ($N = 15$), $y = 1.058x - 2.29 \times 10^{15}$.

The larger scatter at Essex reflects site-specific sampling/representativeness limitations rather than systematic bias in the Pandora retrieval.

- New Figure 8: shows per-layer mean Pandora–aircraft difference ($\pm$ SEM) at RRC2024 ($N = 6$ spirals) and Essex2024 ($N = 11$ spirals). Under the operational mapping, both sites show retrievals exceeding the in situ measurement by roughly 2 ppb below ~ 800 m, with correspondingly lower HCHO abundance between ~ 1000 – 2000 m — the signature of HCHO being over-concentrated into the lowest retrieval layers. Under the revised mapping, the mean difference is close to zero across the full sampled height range at both sites (largest remaining bias: 0.7 ppb, 500–1000 m at RRC).
- Uncertainty-aware column agreement: across all sites, Pandora and aircraft columns agreed within 1σ combined uncertainty for 60% of spiral comparisons (79% excluding Essex), and within 2σ for 100% of comparisons.

3. Data filtering approach: rationale and validation

New manuscript text explains why wrms (normalized fitting-residual RMS) and relative tropospheric-column uncertainty are used in preference to the operational L2 quality flag, which is typically comparably stringent but built for network-wide statistics rather than short-duration campaigns at specific sites. wrms is used as the primary filter because it directly diagnoses fit quality; for this dataset it correlates with solar zenith angle ($r = 0.56$ – 0.58) and with integration time ($r = 0.67$) — the latter also serving as a cloud-contamination proxy, since Pandora's measurement routine automatically lengthens exposure as light levels drop. We note that Rawat et al.'s (2025) data-driven approach to deriving thresholds from the uncertainty distribution is well motivated, but permissive by design.

4. Spatial heterogeneity and wind analysis

We added HRRR ~ 500 m AGL wind fields as a new Figure 3c panel, together with site-specific reasoning for the recurring SW > SE column asymmetry at RRC: the SE viewing sector looks out over vegetated, undeveloped land, while the SW sector points toward the industrialized city of Hopewell, VA. On 2024-06-25, weak morning winds allow the SE/SW asymmetry to persist; by afternoon, winds organize into a moderate south-southwesterly flow and the gradient weakens and reverses sign. Because HCHO is a secondary photochemical product rather than a conserved tracer, its gradient timing is not expected to track wind shifts one-to-one — consistent with what this case shows. For the near-surface HCHO concentration discussion, wind/HCHO panels were added to the SI.

5. Expanded introduction and literature

We added relevant references for Pandora HCHO validation studies (Park et al., 2026; Rawat et al., 2026; Ortega et al., 2026; Chi et al., 2025). We also now mention parameterized MAX-DOAS profile-retrieval approaches alongside optimal estimation, and provide citations supporting the averaging-kernel asymmetry statement.

Reviewer 1 (RC1)

General Assessment

Reviewer: *The paper is relevant, novel, and in scope, but the conclusions — particularly the recommendation to adopt the revised vertical layer mapping across the operational PGN processing — are reached a bit too quickly given the limited dataset, and the results are presented mostly qualitatively. Additional references, discussion of representativeness, and comment on the relationship to Sebol et al. (2026) are requested.*

Response: We thank the reviewer for their comments. The recommendation rests on both a mechanistic explanation for the layer-mapping artifact and the strength of this dataset's spatiotemporal coincidence relative to prior comparisons. These points are now presented more clearly and with more quantitative backing.

Technical and Minor Remarks

Line 37

Reviewer: *Give more references to justify “widely used” for Pandora HCHO satellite validation.*

Response: Citations were added — please see Summary §5.

Line 47

Reviewer: *Parameterized approaches are also common for MAX-DOAS profile retrieval and should be mentioned alongside optimal estimation.*

Response: Addressed, as noted in see Summary §5.

Line 56

Reviewer: *Update the Sebol et al. reference year (2025 → 2026).*

Response: This is now corrected.

Line 84

Reviewer: *Refer to the PGN Level 2 processing version, and comment quantitatively on how much the modified fitting window changes the DSCDs themselves.*

Response: Addressed — see Summary §2 for the full dSCD/O₄ comparison. PGN L2 version number was added at first mention.

Line 125

Reviewer: *Refer to Appendix A when the scaling of partial columns to tropospheric vertical columns is introduced.*

Response: Cross-reference was added.

Line 128

Reviewer: *Provide references supporting the statement that MAX-DOAS sensitivity kernels are strongly asymmetric and bottom-weighted.*

Response: Addressed — please see Summary §5. Citations to optimal-estimation MAX-DOAS papers with representative kernels added.

Lines 156–160

Reviewer: *Why not filter using the operational L2_qa flags? Comment and cite recent papers on limitations of automatic flagging; add column numbers.*

Response: Please see Summary §3 for the full rationale. With expanded explanations of the quantities used for data screening, we decided against referring to column numbers.

Line 171

Reviewer: *Provide a reference or website for the Aeris Ultra in situ sampler, and its uncertainty.*

Response: Website reference and stated precision were added.

Line 182

Reviewer: *Reference Sebol et al. (2026), who found aircraft-measured vs. climatological P/T conversion differ by less than 5%.*

Response: We considered this but chose not to cite it: since we used aircraft-measured P/T directly, we have not independently tested the sensitivity of the climatological assumption at altitude (where HCHO abundance is low) for our own dataset and would rather not cite a result we are not confident about extending here.

Line 200 / Figure 3

Reviewer: *The sampling-geometry sensitivity is interesting given the usual assumption of HCHO spatial homogeneity; assess wind vs. genuine heterogeneity. Same comment for Fig. 5.*

Response: Please see the discussion in Summary §4.

Figure 4

Reviewer: *Add regression statistics (slope, intercept) in addition to the correlation coefficient.*

Response: These are now included (Summary §2).

Line 223

Reviewer: *How long were the collocated Aeris in situ measurements run?*

Response: The Aeris instrument was deployed only on the days shown.

Lines 225, 240

Reviewer: *Provide typical horizontal extent values (as given in the Pandora files) for context.*

Response: Typical horizontal sensitivity extents (~ 11 km across the sites in this study) from the Pandora product files have been added.

Figure 5

Reviewer: *Note in the caption that this is the RRC site, and confirm whether the Pandora error bars are those provided in the Pandora files.*

Response: We updated the caption to identify RRC. The caption now also specifies that the plotted error bars are the independent uncertainty values.

Line 256

Reviewer: *Clarify whether the “operational Pandora layer mapping” panel is the true operational product or the improved DOAS fitting with the unchanged mapping.*

Response: We clarified that the “operational” data are from the true operational rfuh5 product.

Line 261

Reviewer: *The bulge near the surface and rapid drop-off above ~1 km should also be linked to the similar shape problem reported by Sebol et al.*

Response: We added a Sebol citation in this section.

Line 277 (and 283)

Reviewer: “Broadly reproduced” / “the degree of agreement” should be made more quantitative, e.g., partial-column estimates by layer.

Response: As stated in Summary §2, relevant quantitative statistics were added.

Line 296

Reviewer: “Consistent within measurement uncertainty” is not currently demonstrated quantitatively.

Response: Same as above.

Line 305

Reviewer: The recommendation to adopt the revised height mapping network-wide seems quick given the limited dataset.

Response: Please see Summary §1 for our response.

Line 310

Reviewer: “...and other GEO instruments” should also mention LEO validation.

Response: We now simply refer to ‘satellites’ without specifying a viewing geometry.

Line 349

Reviewer: Consider adding “The tropospheric column ‘q’ is estimated...” since q is not introduced earlier.

Response: The tropospheric column is now explicitly denoted ‘q’ at first use (Appendix A).

Reviewer 2 (RC2)

General Assessment

Reviewer: The study is a welcome addition, alongside Sebol et al. (2025/2026), to the literature on Pandora/MAX-DOAS performance relevant to satellite validation. Publication is recommended, with requests to expand context/references, clarify methodology, and extract more quantitative conclusions.

Response: We appreciate the reviewer’s comments and have addressed their input throughout via the summary above.

Major Comments

1. Introduction context and references

Reviewer: Add specific references for TEMPO/GEMS/TROPOMI validation studies and expand on HCHO's motivation.

Response: Specific supporting citations were added.

2. Methods 2.1 — elevation angle selection

Reviewer: Expand on why 75°–60° elevation angles are used, what other angles provide, and cite a supporting study.

Response: In Section 2.1, we now explain the geometric basis for near-horizon vs. near-zenith sensitivity (path length scales roughly as $1/\sin(\text{elevation})$), distinguishing the near-horizon scans that maximize near-surface sensitivity from the 75°/60° reference scans.

3. Methods 2.1.1 — Eqs. 1–3

Reviewer: Provide more detail on Eqs. 1–3, including how $VCD(O_2-O_2)$ and $n_{\text{surface}}(O_2-O_2)$ are obtained.

Response: We added details for these equations. $VCD(O_2-O_2)$ and $n_{\text{surface}}(O_2-O_2)$ are both derived climatologically: the surface density from station pressure/temperature via the ideal gas law, and the vertical column from that surface density times a CIRA climatological effective height.

4. Methods 2.1.2 — filtering and cloud contamination

Reviewer: *What is the relationship between the filtering criteria and cloud contamination? Is there a reference justifying the thresholds?*

Response: Please see Summary §3 for details on the filtering criteria. These are now detailed further in Section 3.1.2.

5. Methods 2.2 — CAFE/Aeris QC and calibration

Reviewer: *Discuss QC and calibration for CAFE and Aeris, given Mouat et al. (2024).*

Response: This is now addressed in Section 2.2 of the manuscript: both instruments were independently calibrated pre-deployment via dynamic dilution of an HCHO standard; CAFE calibrations bracket each flight set (St. Clair et al., 2019) with QC by St. Clair on engineering and fit-quality data, and the Aeris was calibrated pre- and post-2024-campaign with no drift noted in the calibration coefficients. Aeris quality control was performed by Hanisco and Pandey.

6. Figure 3 — missing SW profiles at RRC

Reviewer: *Why are there no SW profiles shown at RRC in Figure 3?*

Response: In 2024, RRC spirals were flown only toward the SE; the two 2025 spirals at this site were one each SE and SW. This explanation is now included in the Figure 3 discussion and in supplemental Figure S7.

7. Figure 4 — subset regression

Reviewer: *Provide a linear fit considering all spirals, and a separate fit excluding spirals whose lowest altitude exceeded 500 m.*

Response: Addressed in Summary §2. The updated Figure 4 shows fit results.

8. Figure 5 — wind information

Reviewer: *If available, include prevailing wind information to help interpret the sampled air mass.*

Response: Supporting wind information was added (Summary §4).

9. Figure 6 and per-layer differences

Reviewer: *Would benefit from quantitative per-layer differences and quantified biases for operational vs. modified retrievals.*

Response: Please see response Summary §2 and Figure 8 in the manuscript.

10. Overall quantitative results and uncertainties

Reviewer: *Provide quantitative overall results and discuss uncertainties; explain dataset-size limitations explicitly.*

Response: Please see response Summary §2 and §1.

11. Conclusions — quantitative summary and operational plans

Reviewer: *Revise conclusions with quantitative results; state plans to operationalize the revised mapping and whether it can be recalculated from existing operational files.*

Response: A quantitative summary, building on figures 4 and 8 was added to the Conclusions. Operational plans are now acknowledged in the manuscript: “These recommendations were formalized by the PGN into an ‘rfuh6’ data version, intended to supersede ‘rfuh5’, that is undergoing network-wide internal evaluation.”

12. Relationship to Sebol et al.

Reviewer: *Place the results of this analysis in the context of the recently published Sebol et al. study.*

Response: Addressed; please see Summary §1.

Minor Comments

Abstract, line 25

Reviewer: *Consider mentioning satellite validation among the stated uses of Pandora HCHO observations.*

Response: This was added.

Introduction, line 38

Reviewer: *Consider citing Zoogman et al. (2017) for TEMPO.*

Response: Added.

Methods, line 84

Reviewer: *Revise the grammar of "...constrain the column abundance in the lower tropospheric."*

Response: Corrected to "...in the lower troposphere."

Methods, line 99

Reviewer: *Is the "NASA real time algorithm" the same one described in Cede (2024)?*

Response: Yes, this is now stated explicitly.

Methods, line 174

Reviewer: *Why was the time-coincidence criterion relaxed to ± 30 minutes for the Essex site?*

Response: Pandora 75 at Essex ran on a standard PGN schedule during ALEGROS 2024 (rather than the customized profile-focused schedules used elsewhere), giving a lower profiling cadence. This is now clarified.

Results and discussion, line 264

Reviewer: *Revise the grammar of "...shows a marked difference between even the modified Pandora profile..."*

Response: Revised.