Reviewer 3

Comment 3.1

This manuscript presents a much-needed investigation of Pandora HCHO products, including both direct-sun and sky-scan retrievals. Given there are limited studies evaluating these data products, particularly in Southeast Asia, the work addresses an important need. The authors analyze observations from five Pandora instruments across this region over several years, comparing variability between direct-sun and sky-scan measurements, as well as against satellite observations from OMI. While the study has the potential to make a valuable contribution, substantial revisions are necessary to address concerns related to the methodology and the interpretation of the results.

Response 3.1

We thank the reviewer for the positive assessment and for recognizing the importance of this study in an understudied region. We also agree that substantial improvements were required in the original manuscript. In response, the manuscript has been extensively revised, including the implementation of an uncertainty-based quality control protocol (Rawat et al., 2025), a rigorous DS–SS intercomparison using temporally matched pairs, and a redesigned satellite–ground collocation framework based on overpass-centered temporal matching. In addition, the analysis has been expanded to include TROPOMI and GEMS, alongside OMI, providing a more comprehensive multi-sensor evaluation.

Revised Text

Section 1 (Introduction – final paragraph)

"In this study, we present a comprehensive evaluation of Pandora HCHO observations across Southeast Asia, explicitly distinguishing between Direct-sun and Sky-scan retrievals and assessing their consistency with multiple satellite products (OMI, TROPOMI, and GEMS). By applying an uncertainty-based quality-control framework and a unified temporal collocation strategy, this work aims to quantify how retrieval geometry, temporal sampling, and spatial representativeness jointly influence satellite–ground agreement in tropical environments."

Comment 3.2

line 24-26: Sky-scan does not report total vertical columns. Only the tropospheric columns.

Response 3.2

We thank the reviewer's comment and agree sky-scan does not essentially report total vertical columns. To avoid confusion, we revised to Pandora Level-2 HCHO columns for clarity.

Revised Text

Abstract

"This study evaluates Pandora Level-2 HCHO columns from five Southeast Asian stations (2021–2024), distinguishing between direct-sun (DS) and sky-scan (SS) observations…"

Comment 3.3

line 81-87: I'm not sure this is true that previous studies don't differentiate between sky-scan and Direct sun modes for HCHO. I'd say it is more that there have been very few studies that look into DS HCHO at all because of known issues. Please reword this section. Same thing on line 121.

Response 3.3

We thank the reviewer for this clarification and agree that the original statement was not appropriately framed. We acknowledge that DS and SS retrievals are generally distinguished in the literature. In the revised manuscript, we have reworded this section to reflect that the key gap is not the lack of differentiation, but rather the limited number of studies that systematically evaluate DS-SS retrievals and their implications for satellite validation, particularly in tropical environments.

Revised Test

Section 1 (Introduction)

"Differences between direct-sun and sky-scan retrievals are primarily associated with sampling characteristics and spatial representativeness. While the two retrieval modes may differ in their effective sensitivity to atmospheric structure, this study focuses on their observational behaviour and consistency rather than explicit vertical sensitivity differences. Direct-sun (DS) and sky-scan (SS) retrievals are often analyzed separately in validation studies due to their differing measurement characteristics. Recent work (e.g., Rawat et al., 2025) has proposed approaches to combine DS and SS observations by accounting for systematic differences in bias and sampling. However, the extent to which these retrieval geometries influence satellite–ground agreement, particularly in terms of spatial-temporal representativeness, remains insufficiently quantified."

Comment 3.4

line 113: Weird that you mention TEMPO and Sentinel-4 but not GEMS for a southeast Asia study?

Response 3.4

We thank the reviewer for this comment and fully agree. In the revised manuscript, GEMS has been explicitly included and discussed alongside TROPOMI and OMI, given its strong relevance for Southeast Asia and its role in the PAN-Asia Pandora network. The text has been revised to reflect the importance of GEMS as a geostationary sensor providing high-temporal-resolution observations over the study region.

Revised Text

Section 1.0 (Introduction)

"Complementing these polar-orbiting sensors, the Geostationary Environment Monitoring Spectrometer (GEMS) offers hourly observations over East and Southeast Asia, enabling improved characterization of diurnal variability and reducing temporal sampling mismatches in satellite–ground comparisons (Bak et al., 2019a, b). The combined use of OMI, TROPOMI, and GEMS therefore provides a comprehensive framework to disentangle the relative roles of spatial resolution, temporal sampling, and retrieval geometry in satellite validation."

Comment 3.5

Table 1: State Altitude above sea level?

Response 3.5

We thank the reviewer for this suggestion. The altitude values in Table 1 have been clarified to explicitly indicate that they represent altitude above mean sea level (m a.s.l.).

Revised Text

Table 1. Summary of Pandora monitoring stations used in this study, including location, altitude, product status, and data availability. Data description: Formaldehyde (HCHO) Level 2, Version: rfus5p1-8 and rfuh5p1-8 (Last accessed: 27 Feb 2025).

Station ID	**Station Name**	**Lat**	**Lon**	**Altitude (m a.s.l.)**	**rfus5p1-8 and rfuh5p1-8**		
					Product Status	**Data Start**	**Last Updated**
190s1	Bangkok	13.7847	100.5400	60	Official	20210520	20250221
210s1	Bandung	-6.8948	107.5865	752	Official	20230611	20240920
211s1	Agam	-0.2046	100.3195	865	Official	20220913	20240521
212s1	Pontianak	0.0415	109.3366	1	Official	20240309	20250226
77s1	Singapore	1.2990	103.7710	77	Official	20230621	20250226

Comment 3.6

line 142: sky-scan does not report the same product as DS. Sky-scan reports 'Tropospheric' column (usually 3-4 km) while DS reports total column.

Response 3.6

We thank the reviewer for this important clarification. We agree that sky-scan (SS) and direct-sun (DS) retrievals do not represent identical quantities. DS retrievals provide total column HCHO along the direct solar beam, whereas SS retrievals are generally more sensitive to the tropospheric column and may not fully capture the total column, depending on the retrieval configuration and atmospheric conditions. In the revised manuscript, we have clarified this distinction to avoid ambiguity and to ensure that the differences between DS and SS retrievals are interpreted appropriately.

Revised Text

Section 2.1 (Pandora Data Description)

"We use Level 2 HCHO products from the rfus5p1-8 and rfuh5p1-8 processing version (last accessed: 27 February 2025), which provides HCHO columns derived from direct-sun and diffuse-sky measurements. Direct-sun (DS) retrievals provide total column HCHO along the solar beam, whereas sky-scan (SS) retrievals represent a tropospheric column derived from multi-angle scattered radiation measurements, with sensitivity that depends on retrieval configuration and atmospheric conditions."

Comment 3.7

line 155: I suggest you reword this sentence because Pandora data quality flags are already confusing. All pandoras belong to the official pandora global network, but this does nothing to guarantee the quality of the data.

Response 3.7

We thank the reviewer for this important clarification. We agree that affiliation with the Pandora Global Network does not inherently guarantee data quality, and that the original wording could be misleading. In the revised manuscript, this statement has been reworded to avoid any implication of automatic data reliability. Instead, we emphasize that data quality is ensured through the application of uncertainty-based quality control criteria within this study.

Revised Text

Section 2.1 (Pandora Data Description)

"All Pandora instruments used in this study are part of the Pandora Global Network; however, data quality is not assumed a priori and is evaluated using uncertainty-based quality control criteria applied in this work."

Comment 3.8

line173-175: You don't need to include the file versions in the results. The methods is sufficient.

Response 3.8

We thank the reviewer for this suggestion and agree that including file version details in the Results section is unnecessary. In the revised manuscript, all dataset version information has been removed from the Results section and retained only in the Methods section, where it is more appropriate.

Comment 3.9

Table 2: move to section 2.1

Response 3.9

We thank the reviewer for this suggestion. We agree that tables describing data characteristics are typically placed in the Methods section. However, in the revised manuscript, Table 2 has been substantially updated to present comparative statistics before and after the application of uncertainty-based quality control (QC). As such, Table 2 now reflects the impact of QC on the dataset, including changes in data distribution and retained observations, and is therefore an integral part of the results analysis rather than a data description. For this reason, we have retained Table 2 in the Results section.

Revised Text

Section 3.1 (Results – reference to Table 2)

"As summarised in Table 2, the QC protocol affects direct-sun (DS) and sky-scan (SS) retrievals differently. DS observations generally exhibit higher intrinsic stability, with only modest reductions in retained measurements at most stations (≤3 % at Bangkok and Pontianak)."

Table 2. Summary of Pandora Level-2 formaldehyde (HCHO) observations from Direct-sun (DS) and Sky-scan (SS) retrievals at selected Southeast Asian stations. Data are categorized by quality flags into High (0,10), Medium (1,11), Low (2,12), and Unusable (≥20). The Rawat quality control (QC) protocol was applied to filter observations based on independent uncertainty thresholds, relative uncertainty (<10%), WRMS (<0.01), and maximum horizontal distance (MHxD < 20 km for SS). Totals represent the number of valid matched DS–SS observation pairs used in the analysis.

Station	Dataset	High	Medium	Low	Unusable	Total
Bangkok	DS Raw	11,693	7,339	61,304	0	80,336
	SS Raw	65,921	35,305	34,438	0	135,664
	DS After QC	11,692	7,301	60,079	0	79,072
	SS After QC	51,119	19,389	8,564	0	79,072
Bandung	DS Raw	8,995	2,113	23,140	0	34,248
	SS Raw	32,766	6,223	8,172	0	47,161
	DS After QC	8,321	1,896	20,051	0	30,268
	SS After QC	26,447	2,448	1,373	0	30,268
Agam	DS Raw	5,438	849	20,375	8,842	35,504
	SS Raw	2,450	445	37,545	0	40,440
	DS After QC	4,582	627	13,413	182	18,804
	SS After QC	1,623	165	17,016	0	18,804
Pontianak	DS Raw	6,779	1,015	17,900	0	25,694
	SS Raw	27,300	1,681	7,298	0	36,279
	DS After QC	6,696	996	17,403	0	25,095
	SS After QC	22,204	443	2,448	0	25,095
Singapore-NUS	DS Raw	5,942	1,771	32,078	0	39,791
	SS Raw	18,633	18,000	24,840	0	61,473
	DS After QC	5,209	1,496	25,750	0	32,455
	SS After QC	13,623	11,279	7,553	0	32,455

Comment 3.10

Table 2: While explaining the data quality is necessary for Pandora discussions, I think there is more relevant information that would explain the Pandoras better. Because basically all data is unassured it doesn't do us much good to focus on that. Instead, I would rather see the 'high' 'medium' and 'low' flags that are also included in the L2 files.

Response 3.10

We thank the reviewer for this helpful suggestion and agree that the high-, medium-, and low-quality classifications provide more meaningful insight than the broader assured/not-assured grouping. In the revised manuscript, Table 2 has been updated to explicitly present the distribution of high- (QF = 0, 10), medium- (QF = 1, 11), and low-quality (QF = 2, 12) retrievals, both before and after the application of uncertainty-based QC. This provides a clearer representation of data quality characteristics and the impact of QC filtering across stations. Please see Response 3.9 for the updated Table 2.

Comment 3.11

The DS HCHO is all unassured because the PGN does not have an official method of assuring that product (I'm not sure what is going on with Singapore, but if you have not I suggest reaching out to the operator to make sure the assured values are real).

Response 3.11

We thank the reviewer for this important clarification regarding the assured status of Pandora HCHO retrievals. We acknowledge that DS HCHO products are often flagged as "unassured" within the PGN framework, reflecting the absence of an official assurance procedure rather than necessarily indicating poor data quality. In the revised manuscript, we do not rely on the assured/not-assured classification to determine data quality. Instead, we apply an uncertainty-based quality control framework (Rawat et al., 2025), which evaluates each observation based on independent criteria such as relative uncertainty, spectral fitting residual (WRMS), and spatial representativeness constraints. This approach ensures that data quality is assessed consistently across all sites, regardless of PGN assurance status. To avoid confusion, the manuscript has been revised to clarify the interpretation of Pandora quality flags and to emphasize that data selection is based on quantitative QC metrics rather than PGN assurance categories.

Revised Text

Section 2.1.1 (Quality Flag Clarification)

"In this way, the assured/not-assured classification within the Pandora Global Network does not directly determine data usability for this study. Instead, data quality is evaluated using uncertainty-based criteria, including relative uncertainty, spectral fitting residual (WRMS), and additional screening parameters, ensuring consistent selection of physically reliable observations."

Comment 3.12

Once the backlog of manually assuring the data is complete, much of the Sky-scan data should be fine, however this table does not show that. I am surprised that there is no assured data at all for any of these sites.

Response 3.12

We acknowledge that the absence of assured data in Table 2 may appear unexpected, particularly given that sky-scan retrievals are generally considered reliable once formally reviewed within the PGN framework. However, in this study, the analysis does not rely on the assured/not-assured classification, as this status reflects the administrative validation process within the PGN, which may not yet be completed for all datasets. Instead, data quality is evaluated using an uncertainty-based quality control framework (Rawat et al., 2025), which applies consistent and quantitative criteria across all observations. The lack of 'assured' classification for DS retrievals reflects current PGN processing status rather than data invalidity and does not preclude their use when uncertainty-based criteria are satisfied. As a result, Table 2 reflects the raw PGN quality flag distribution, while the actual data selection for analysis is based on independent QC metrics. This approach ensures that high-quality observations are retained regardless of their formal assurance status.

Revised Text

Section 2.1.1 (QC Protocol)

To improve the robustness of ground-based HCHO observations used for intercomparison and satellite validation, an uncertainty-based quality control (QC) protocol following the methodological framework of Rawat et al. (2025) was applied to contemporaneous Pandora direct-sun (DS) and sky-scan (SS) observations. DS and SS retrievals were first paired within a 5 min tolerance window. A high-quality reference subset was then defined using Pandora quality flags QF = 0 or 10 for both DS and SS retrievals, and dynamic absolute uncertainty thresholds were calculated separately for DS and SS as the mean plus three standard deviations of the uncertainty in this subset. Matched observations were retained when either both DS and SS absolute uncertainties were below these dynamic thresholds or both relative uncertainties were below 10 %. Additional filters required WRMS < 0.01 for both DS and SS retrievals and, for sky-scan observations, maximum horizontal distance (MHxD) < 20 km when available. Pandora quality flags were subsequently used to classify observations into high-quality (QF = 0, 10), medium-quality (QF = 1, 11), low-quality (QF = 2, 12), and unusable (QF $\geq$ 20) categories for diagnostic analysis. This procedure reduces the influence of retrieval noise, poor spectral fits, and unfavorable viewing geometry prior to satellite collocation.

Comment 3.13

Are these data only for the OMI overpass time?

Response 3.13

We thank the reviewer for this question. In the revised manuscript, the data are not limited to the exact OMI overpass time. Pandora observations are collocated with satellite measurements using two approaches: nearest-time matching (±2 h) and overpass-centered averaging windows (±30 min, ±1 h, ±2 h). This allows evaluation of temporal representativeness while ensuring consistency with satellite sampling. All analyses are based on temporally collocated observations rather than fixed overpass-only sampling.

Revised Text

Section 2.3 (Collocation Strategy)

"Two complementary approaches were applied. First, a nearest-time matching method paired each satellite observation with the closest Pandora measurement within a ±2 h tolerance window. Second, an overpass-window averaging method was used, in which all Pandora observations within symmetric windows centered on the satellite overpass time were averaged to form representative ground-based column estimates. Three temporal windows were tested (±30 min, ±1 h, and ±2 h) to assess sensitivity to temporal smoothing."

Comment 3.14

Line 188: Reword

Response 3.14

This line has been removed due to the substantial revision in the revised manuscript.

Comment 3.15

Line 189-197: No need to type out everything that is already in Table 2.

Response 3.15

These lines are now removed, partly due to the substantial revision in the revised manuscript.

Comment 3.16

Line 190: I think you should rethink your exclusion criteria. There is no official recommendation for excluding bad quality data, however several studies suggest removing data based on uncertainty (Rawat et al 2024).

Response 3.16

We thank the reviewer for this important suggestion and fully agree that uncertainty-based filtering provides a more robust approach than relying solely on Pandora quality flags. In the revised manuscript, we have implemented an uncertainty-based quality control protocol following Rawat et al. (2025). This includes filtering based on relative uncertainty (<10%), spectral fitting residual (WRMS < 0.01), and additional criteria such as MHxD for sky-scan retrievals, ensuring that only physically reliable observations are retained.

Revised Text

Section 3.1 (Implications of Rawat QC Protocol)

"The application of the uncertainty-based quality-control (QC) protocol following Rawat et al. (2025) substantially improves the statistical robustness and physical representativeness of Pandora Level-2 HCHO retrievals across the Southeast Asian network. The QC procedure integrates formal quality flags with independent uncertainty metrics, including relative uncertainty (<10 %), spectral fitting residual (WRMS < 0.01), and spatial representativeness constraints for sky-scan (SS) observations. The filtering primarily targets retrieval artefacts associated with viewing-geometry sensitivity and low signal-to-noise conditions, while preserving the underlying atmospheric variability."

Comment 3.17

Table3: Don't need version numbers in description.

Response 3.17

Revised. The version number is now removed in Table 3.

Table 3. Descriptive statistics of Pandora Level-2 formaldehyde (HCHO) retrieved from Direct-sun (DS) and Sky-scan (SS) observations at selected Southeast Asian stations. Statistics are shown for contemporaneous matched DS–SS pairs before and after applying the Rawat quality control (QC) protocol. Values are reported as mean ± standard deviation (SD), median with interquartile range (IQR; Q1–Q3), and minimum–maximum. All HCHO columns are expressed in units of $\times 10^{15}$ molecules cm^{-2}.

Station	Dataset	Mean ± SD	Median (IQR)	Min–Max	N
(a) Direct-sun (DS) HCHO					
Bangkok	Raw	21.2 ± 7.88	20.6 (16.0–25.7)	0.018–200	80,336
	After QC	21.3 ± 7.83	20.7 (16.1–25.8)	0.018–83.6	79,072
Bandung	Raw	19.6 ± 44.4	15.0 (10.2–21.0)	0.012–988	34,248
	After QC	16.4 ± 8.29	15.3 (10.5–21.1)	0.012–76.9	30,268
Agam	Raw	63.0 ± 166.6	9.93 (6.76–14.0)	0.013–999	35,504
	After QC	9.01 ± 4.07	8.93 (6.40–11.2)	0.013–77.0	18,804
Pontianak	Raw	11.9 ± 7.90	11.7 (8.93–14.4)	0.013–569	25,694
	After QC	11.8 ± 5.17	11.7 (9.00–14.4)	0.013–107	25,095

Station	Dataset	Mean ± SD	Median (IQR)	Min–Max	N
Singapore-NUS	Raw	10.8 ± 6.89	9.33 (6.53–13.2)	0.012–131	39,791
	After QC	10.9 ± 7.23	9.30 (6.36–13.3)	0.012–131	32,455
(b) Sky-scan (SS) HCHO					
Bangkok	Raw	12.8 ± 10.8	11.9 (8.31–16.1)	0.013–956	135,664
	After QC	13.2 ± 5.57	12.6 (9.43–16.3)	0.025–61.3	79,072
Bandung	Raw	13.0 ± 17.6	11.0 (6.67–16.6)	0.014–923	47,161
	After QC	13.1 ± 7.18	11.8 (7.91–17.1)	0.026–116	30,268
Agam	Raw	7.40 ± 26.9	4.39 (2.52–6.78)	0.010–992	40,440
	After QC	4.93 ± 3.07	4.51 (2.92–6.40)	0.010–54.5	18,804
Pontianak	Raw	8.10 ± 15.0	6.75 (4.46–9.77)	0.010–919	36,279
	After QC	8.22 ± 3.82	7.62 (5.53–10.2)	0.027–45.2	25,095
Singapore-NUS	Raw	10.9 ± 11.1	9.02 (6.05–13.4)	0.013–883	61,473
	After QC	11.7 ± 7.34	9.90 (7.11–14.0)	0.024–90.6	32,455

Comment 3.18

Figure 2: What are the colors of the bars? If nothing, make the bars the same color.

Response 3.18

In the revised manuscript, Figure 2 is removed and replaced with Figure 6.

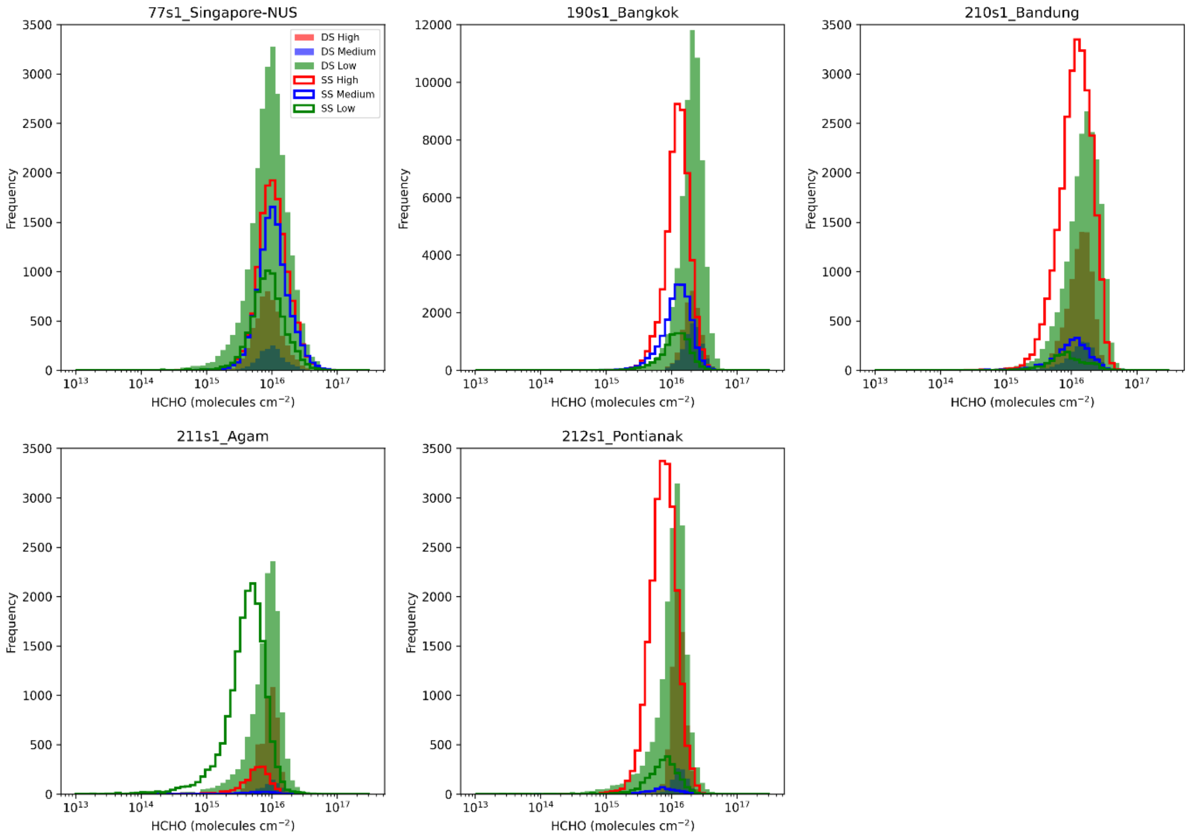

Figure 6. Frequency distributions of filtered HCHO vertical column densities (molecules cm^{-2}) for direct-sun (DS; shaded histograms) and sky-scan (SS; solid line histograms) observations across five stations. Each panel corresponds to a station, with HCHO data grouped by quality flag: high quality (red; QF = 0, 10), medium quality (blue; QF = 1, 11), and low quality (green; QF = 2, 12). The x-axis is shown on a logarithmic scale to capture the wide dynamic range of HCHO values. The y-axis represents the frequency of observations, with consistent limits applied across stations except for Bangkok, which has a higher observation count.

Comment 3.19

Figures 2 and 3: I suggest changing these to a normalized distribution plot. Because some Pandoras have more data than others we would be able to see the differences better.

Comment 3.20

If the message is to compare DS to Sky-scan I also suggest including both results on the same subplot. For example a histogram of the one monitor's DS values in one color and the Sky-scan values in another color on top of that histogram. As it is now, it is hard to compare.

Response 3.19 & 3.20

We thank the reviewer for this helpful suggestion and agree that direct comparison between DS and sky-scan retrievals should be clearly presented. In the revised manuscript, the primary analysis of DS–SS consistency has been updated to use scatter plot (correlation) analysis of temporally matched pairs, which provides a more direct and quantitative assessment of agreement between retrieval modes. As the focus of this study is on evaluating consistency and variability between DS and SS observations, correlation-based diagnostics are considered more appropriate.

Revised Text

Section 3.2 (DS–SS Comparison)

The nine-panel correlation analyses (Figs. 3) reveal that DS–SS agreement depends strongly on retrieval quality category, with the highest correlations observed when both measurements fall within the high-quality regime (QF = 0, 10). Across all stations, high-quality DS–SS pairs exhibit correlation coefficients typically exceeding 0.70, indicating strong agreement between retrieval geometries under well-constrained uncertainty conditions.

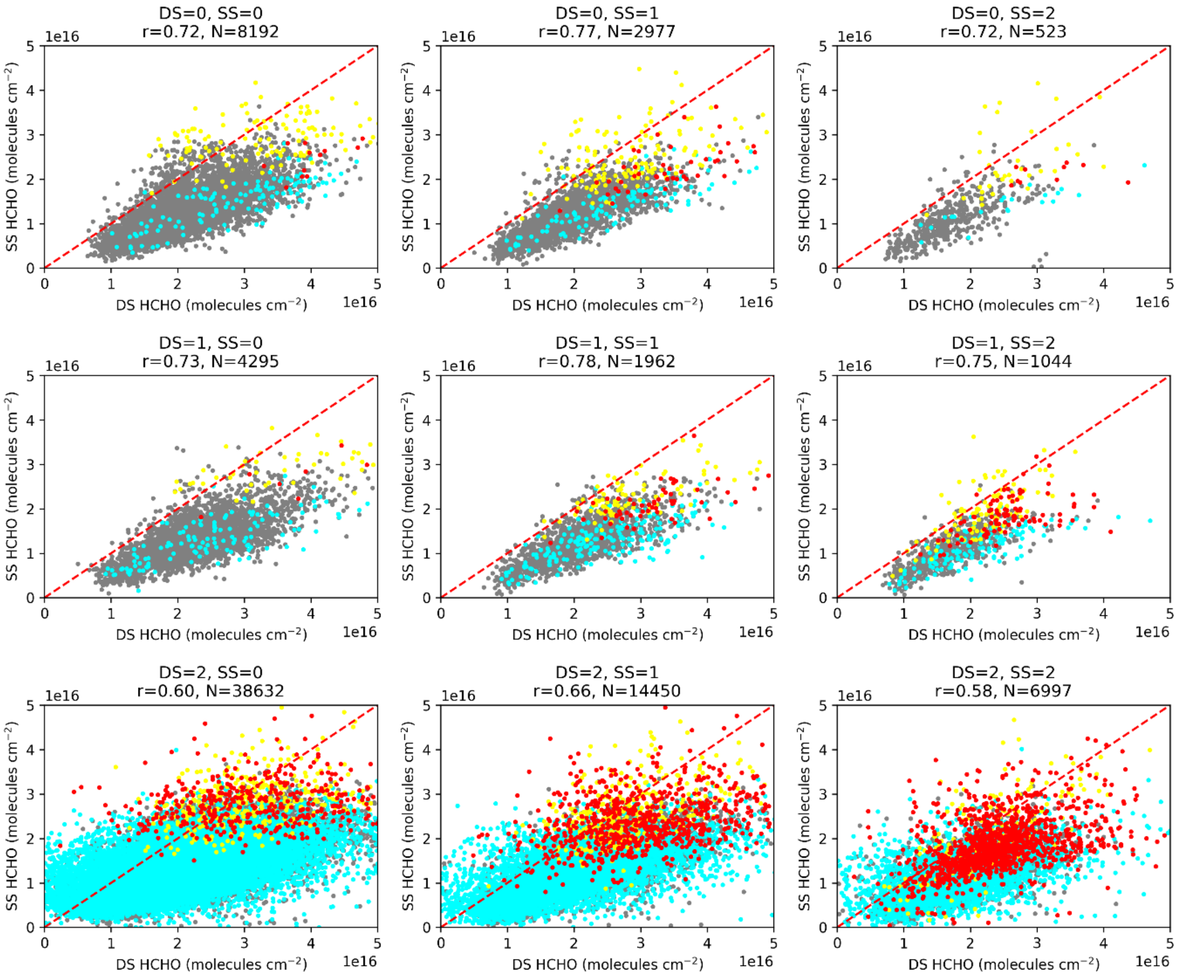

Figure 3. Nine-panel plot of correlation between contemporaneous Pandora HCHO column amounts: direct-sun (DS) vs sky-scan (SS) for each quality category, following the Rawat et al. (2025, AMT) QC method at Bangkok station. Panels are organized by DS and SS quality categories (0 = high, 1 = medium, 2 = low). Each panel shows the scatter of DS vs SS HCHO (molecules cm^{-2}), with points color-coded by uncertainty thresholds: gray = both below cutoff, cyan = DS above cutoff, yellow = SS above cutoff, red = both above cutoff. The red dashed line represents the 1:1 relationship, and the correlation coefficient (r) and number of matched observations (N) are indicated in each panel.

Comment 3.21

Figure 4: In this figure description you change your filtering methods to also remove data above 50e15. This needs to be consistent throughout the entire results and stated in methodology. What is the reasoning behind this number? Instead I suggest filtering based on uncertainty and that would most likely give a similar result.

Response 3.21

We thank the reviewer for this important comment and fully agree that applying an arbitrary fixed threshold (e.g., 50 × 10^{15} molecules cm^{-2}) is not appropriate without consistent justification. In the revised manuscript, this threshold-based filtering has been removed entirely. All data screening is now performed using an uncertainty-based quality control framework following Rawat et al. (2025), including criteria based on relative uncertainty, spectral fitting residual (WRMS), and additional quality constraints. This ensures a physically consistent and objective selection of valid observations. The QC approach is now applied consistently across all analyses, and no additional ad hoc thresholds are used.

Revised Text

Section 2.1.1 (QC Protocol)

"To improve the robustness of ground-based HCHO observations used for intercomparison and satellite validation, an uncertainty-based quality control (QC) protocol following the methodological framework of Rawat et al. (2025) was applied to contemporaneous Pandora direct-sun (DS) and sky-scan (SS) observations. DS and SS retrievals were first paired within a 5 min tolerance window. A high-quality reference subset was then defined using Pandora quality flags QF = 0 or 10 for both DS and SS retrievals, and dynamic absolute uncertainty thresholds were calculated separately for DS and SS as the mean plus three standard deviations of the uncertainty in this subset. Matched observations were retained when either both DS and SS absolute uncertainties were below these dynamic thresholds or both relative uncertainties were below 10 %. Additional filters required WRMS < 0.01 for both DS and SS retrievals and, for sky-scan observations, maximum horizontal distance (MHxD) < 20 km when available."

Comment 3.22

You reuse letters in labeling the panels in this figure (and Figs 2-3). Each panel needs a unique letter. I also suggest combining the hourly and daily figures together so we can easily see the differences.

Response 3.22

The panel labeling has been revised so that each subplot is assigned a unique identifier, ensuring clarity and consistency across all figures in the revised manuscript. Regarding the suggestion to combine hourly and daily plots, these figures have been removed, and the result visualization have been reorganized in the revised manuscript to improve readability and to better align with the key analysis objectives. Rather than combining multiple temporal aggregations, the revised figures focus on different collocation strategies, which present the most relevant information clearly.

Revised Text

Section 3.4 (Impacts of Collocation Strategy)

"The impact of temporal collocation strategy differs markedly between OMI and TROPOMI. For OMI (Fig. 8), expanding the collocation window from nearest to ±2 h results in only marginal changes in correlation and error metrics, indicating limited sensitivity to temporal averaging. For instance, at Bangkok, correlation remains nearly unchanged ($r = 0.11–0.12$), while RMSE ($\sim 1.13–1.16 \times 10^{16}$ molecules cm^{-2}) and MAE ($\sim 9.0–9.9 \times 10^{15}$ molecules cm^{-2}) show minimal improvement. In contrast, TROPOMI (Fig. 11) demonstrates clearer benefits from temporal averaging. At Bangkok, the correlation increases from $r = 0.23$ (nearest) to $r = 0.32$ (±2 h), accompanied by a reduction in RMSE from 9.35×10^{15} to 8.80×10^{15} molecules cm^{-2} and a decrease in MAE from 7.47×10^{15} to 7.07×10^{15} molecules cm^{-2}. A similar but more subtle improvement is observed at Singapore, where the correlation remains consistently high ($r \approx 0.46$–0.47), while RMSE decreases from 8.34×10^{15} to 6.47×10^{15} molecules cm^{-2} and MAE from 5.42×10^{15} to 4.81×10^{15} molecules cm^{-2} when applying a ±2 h window. These results indicate that TROPOMI retrievals benefit from temporal averaging while maintaining strong correlation, reflecting improved representation of short-timescale variability compared to OMI."

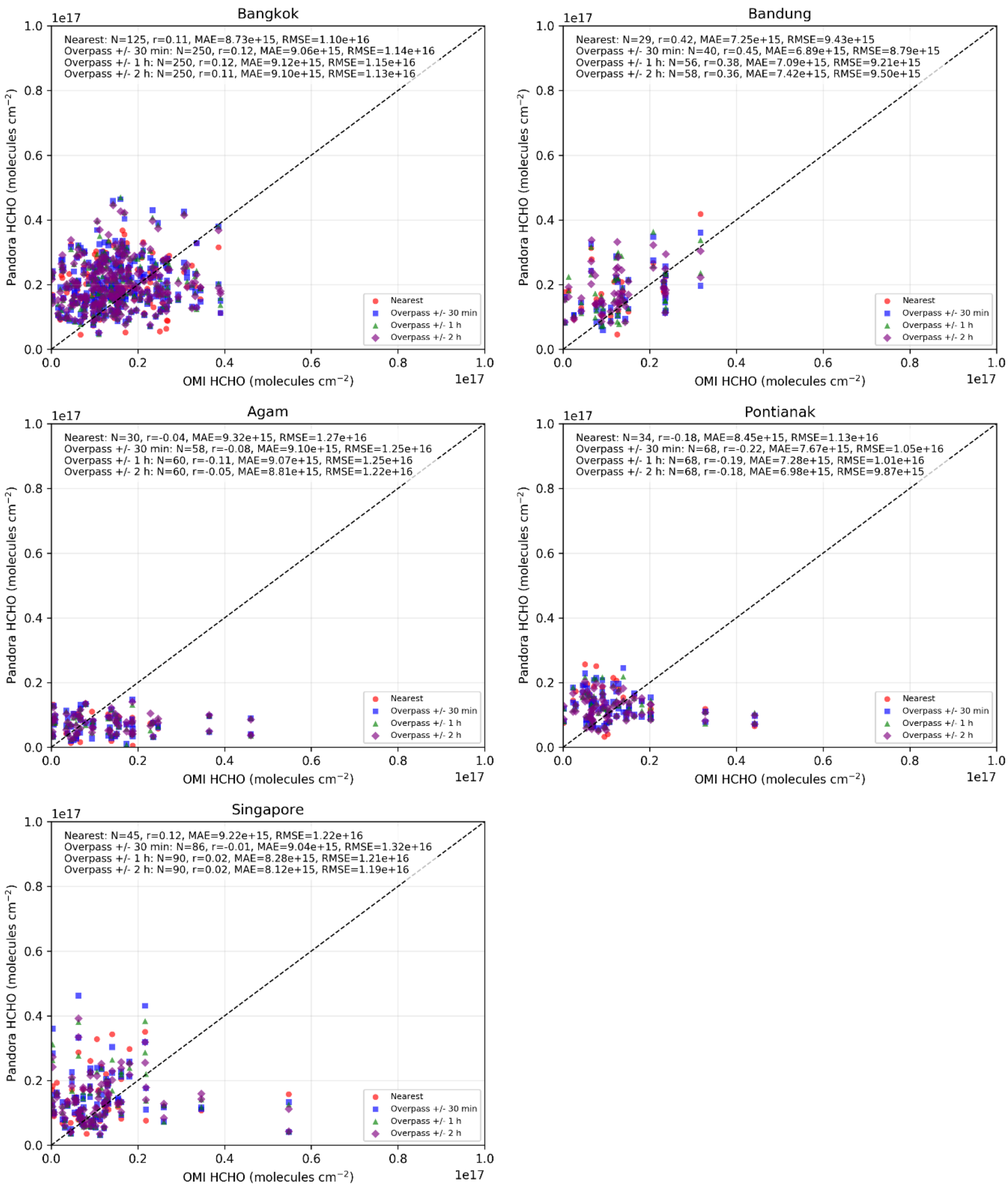

Figure 8. Scatter plots comparing Pandora and OMI HCHO column retrievals for different temporal collocation strategies. Each panel corresponds to a measurement station and includes the 1:1 reference line. Reported statistics include sample size (N), mean absolute error (MAE), root-mean-square error (RMSE), and Pearson correlation coefficient (r).

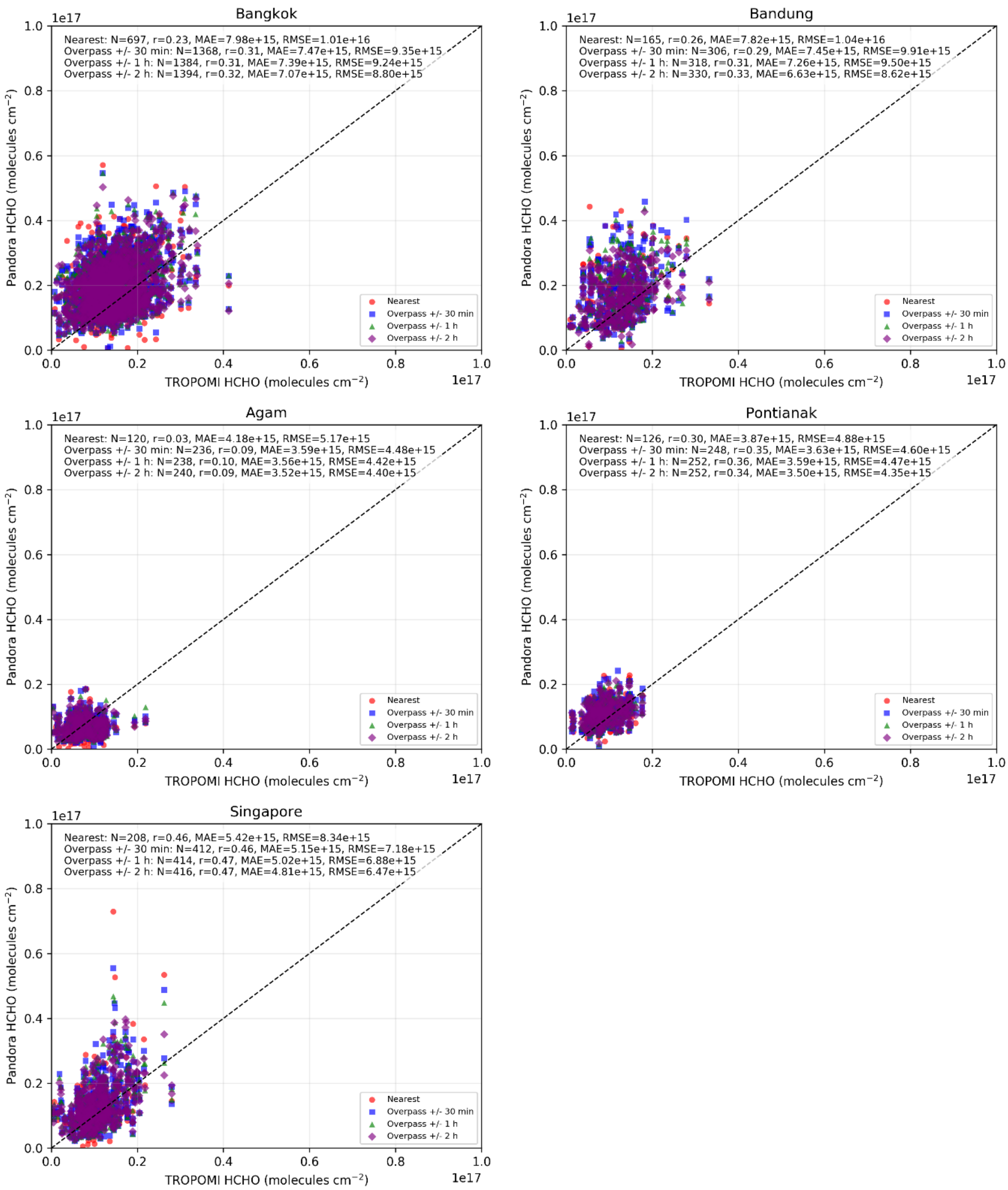

Figure 11. Scatter plots comparing Pandora and TROPOMI HCHO column retrievals for different temporal collocation strategies. Each panel corresponds to a measurement station and includes the 1:1 reference line. Reported statistics include sample size (N), mean absolute error (MAE), root-mean-square error (RMSE), and Pearson correlation coefficient (r).

Comment 3.23

Table 4: Under "remarks" are the distances needed? I don't understand what that is referring to other than the OMI spatial averaging column.

Response 3.23

In the revised manuscript, Table 4 and the associated experimental framework have been removed as part of the redesign of the collocation methodology. The analysis now uses a unified time-based collocation approach, and the ambiguity associated with the previous table has been eliminated.

Comment 3.24

Why noontime if the OMI overpass is closer to 1pm? What are the time windows for "daytime". Pandora does not report data at nighttime.

Response 3.24

We thank the reviewer for this comment and fully agree that fixed time windows such as "noontime" or "daytime" are not appropriate for satellite validation, particularly given the well-defined OMI overpass time and the strong diurnal variability of HCHO. In the revised manuscript, this framework has been removed entirely. Pandora observations are now collocated with satellite data using overpass-centered temporal matching, including nearest-time pairing (±2 h) and symmetric averaging windows (±30 min, ±1 h, ±2 h) around the satellite overpass time. This approach ensures temporal consistency and avoids ambiguity associated with fixed time definitions.

Revised Text

Section 2.3 (Collocation Straetgy)

"Two complementary approaches were applied. First, a nearest-time matching method paired each satellite observation with the closest Pandora measurement within a ±2 h tolerance window. Second, an overpass-window averaging method was used, in which all Pandora observations within symmetric windows centered on the satellite overpass time were averaged to form representative ground-based column estimates. Three temporal windows were tested (±30 min, ±1 h, and ±2 h) to assess sensitivity to temporal smoothing."

Comment 3.25

Figure 5: x tick times need to be more clear. State month.

Response 3.25

The x-axis labels in Figure 7 and Figure 10 (after revised) have been updated to explicitly include month information, improving clarity and readability of the time series.

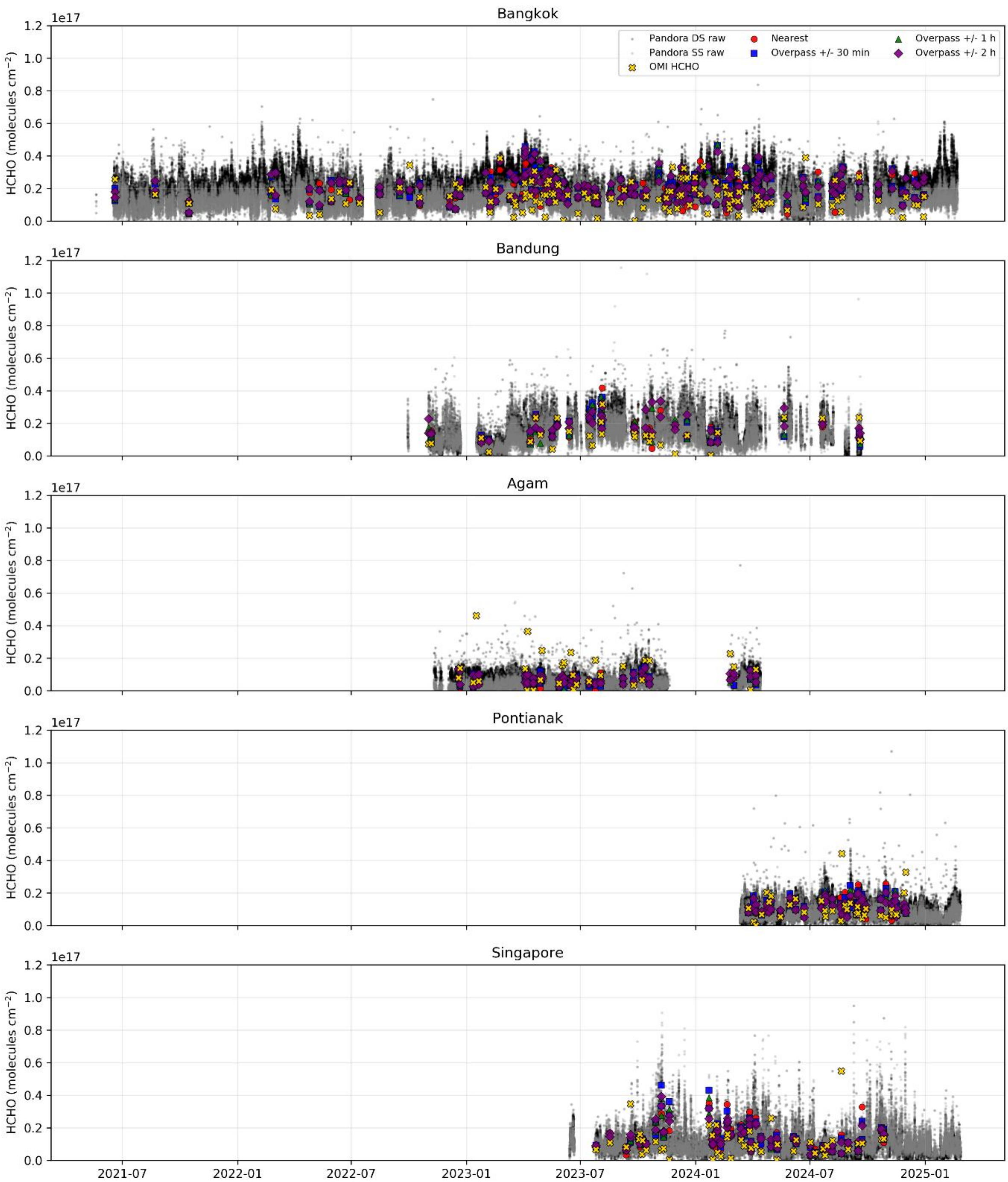

Figure 7. Time series of Pandora HCHO column measurements (DS and SS) and temporally collocated OMI observations at five Southeast Asian stations. OMI-Pandora data are shown for four collocation approaches: nearest-time matching and overpass-centred averaging windows of ±30 min, ±1 h, and ±2 h.

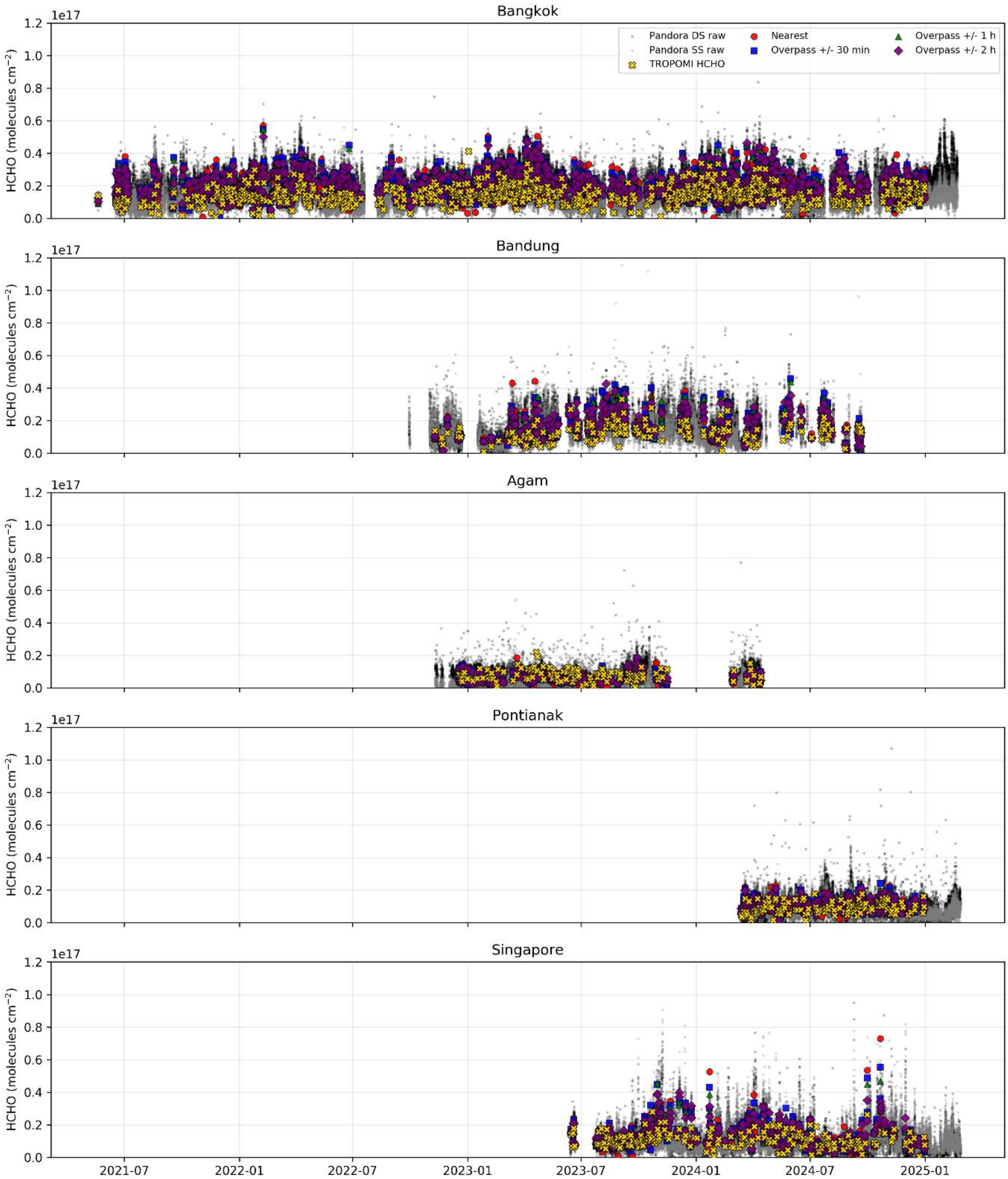

Figure 10. Time series of Pandora HCHO column measurements (DS and SS) and temporally collocated TROPOMI observations at five Southeast Asian stations. TROPOMI-Pandora data are shown for four collocation approaches: nearest-time matching and overpass-centred averaging windows of ±30 min, ±1 h, and ±2 h.

Comment 3.26

Reword the description. The different experiments E1-E9 are referring to OMI not the Pandoras, right?

Response 3.26

We thank the reviewer for this comment and agree that the description of the E1–E9 experiments was unclear. In the revised manuscript, the entire E1–E9 experimental framework has been removed and replaced with a unified, physically consistent collocation methodology. As a result, this ambiguity no longer arises in the revised version.

Comment 3.27

I think you are being a little hasty in determining which experiment is performing best. This needs more discussion in the text. In Table 5 there does not seem to be a clear winner, rather each Pandora monitor works best with a different comparison method.

Response 3.27

We thank the reviewer for this important comment and agree that the identification of a single "best" experiment in the previous analysis was not sufficiently robust. In the revised manuscript, the E1–E9 experimental framework and Table 5 have been removed and replaced with a unified collocation approach based on overpass-centered temporal matching. As a result, the comparison is no longer framed in terms of selecting a single optimal method. Instead, the revised discussion emphasizes how satellite–ground agreement depends on temporal window selection and site-specific conditions, without implying a universal best configuration. This provides a more physically consistent and generalizable interpretation.

Revised Text

Section 3.4 (Impact of Collocation Strategy)

"The impact of temporal collocation strategy differs markedly between OMI and TROPOMI. For OMI (Fig. 8), expanding the collocation window from nearest to ±2 h results in only marginal changes in correlation and error metrics, indicating limited sensitivity to temporal averaging. For instance, at Bangkok, correlation remains nearly unchanged ($r = 0.11$–0.12), while RMSE (~1.13–1.16×10^{16} molecules cm^{-2}) and MAE (~9.0–9.9×10^{15} molecules cm^{-2}) show minimal improvement. In contrast, TROPOMI (Fig. 11) demonstrates clearer benefits from temporal averaging. At Bangkok, the correlation increases from $r = 0.23$ (nearest) to $r = 0.32$ (±2 h), accompanied by a reduction in RMSE from 9.35×10^{15} to 8.80×10^{15} molecules cm^{-2} and a decrease in MAE from 7.47×10^{15} to 7.07×10^{15} molecules cm^{-2}. A similar but more subtle improvement is observed at Singapore, where the correlation remains consistently high ($r \approx 0.46$–0.47), while RMSE decreases from 8.34×10^{15} to 6.47×10^{15} molecules cm^{-2} and MAE from 5.42×10^{15} to 4.81×10^{15} molecules cm^{-2} when applying a ±2 h window.

These results indicate that TROPOMI retrievals benefit from temporal averaging while maintaining strong correlation, reflecting improved representation of short-timescale variability compared to OMI."

Comment 3.28

Figure 6: No need for different colors in this figure.

Response 3.28

This figure is now removed in the revised manuscript.

Comment 3.29

Line 315-319: This figure needs to be discussed in greater detail. I think this could be an important figure to show the comparison between OMI and Pandora HCHO distributions.

Response 3.29

We thank the reviewer for this helpful suggestion and agree that this figure provides important insight into the comparison between Pandora and satellite HCHO observations. In the revised manuscript, the discussion has been substantially expanded. For the DS–SS comparison, we provide a more detailed interpretation of the distributional characteristics, highlighting how differences in distribution shape and spread reflect retrieval quality and sampling-related variability.

For the satellite comparison, the analysis has been fully redesigned to provide a more comprehensive and physically grounded evaluation. Specifically, we now assess Pandora–satellite consistency using multiple complementary diagnostics, including:

(i) time-series comparison,

(ii) correlation analysis (DS and SS combined),

(iii) bias as a function of Pandora short-timescale variability,

(iv) separate correlation analysis for DS and SS retrievals.

This multi-dimensional framework provides a more complete interpretation of satellite–ground agreement and allows clearer attribution of discrepancies to temporal sampling and spatial representativeness effects.

Revised Text

Section 3.3 (Distributional characteristics DS-SS retrievals)

"The frequency distributions highlight important station-dependent differences in DS–SS consistency. At the Pandora Singapore-NUS station, DS and SS distributions largely overlap across all QF categories, indicating strong consistency between retrieval geometries irrespective of quality classification. However, this behaviour is not observed at the other stations. At Bangkok, Bandung, Pontianak, and Agam, SS retrievals are systematically skewed towards lower HCHO values compared to DS, particularly in the medium- and low-quality regimes. This systematic shift suggests a geometry-dependent bias, where SS measurements tend to underestimate column HCHO relative to DS under less favourable retrieval conditions. Such discrepancies are reduced in the high-quality category but remain evident overall, highlighting the importance of quality filtering when combining DS and SS observations for quantitative analysis."

Section 3.4 Impact of temporal collocation and sub-pixel variability on OMI and TROPOMI validation

Section 3.5 Role of High-Temporal-Resolution Observations: Insights from GEMS HCHO Retrievals

Comment 3.30

Figures 7and 8: combine into one showing E1 in one color, and E2/E8 in another for easier comparison. Are these the daily comparisons? Noontime?

Response 3.30

We thank the reviewer for this suggestion and agree that the previous figure design and labeling were not sufficiently clear. In the revised manuscript, Figures 7 and 8 and the associated E1–E9 experimental framework have been removed as part of the redesign of the analysis. The satellite–Pandora comparison is now presented using a unified and physically consistent approach, including time-series comparison, correlation analysis, and overpass-centered temporal matching, which avoids the ambiguity associated with the previous experiment-based figures. As a result, the issues related to figure combination and unclear temporal definitions (e.g., daily vs. noontime) no longer arise in the revised version.

Comment 3.31

lines 320-325: More discussion on this figure as well. This seems to be the main point of the paper yet only a few sentences.

Response 3.31

We thank the reviewer for this important comment and agree that this figure represents a central result of the study and requires more detailed discussion. In the revised manuscript, the interpretation of this figure has been substantially expanded to provide a clearer and more comprehensive explanation of the differences between DS and SS retrievals in the context of satellite comparison. The discussion now explicitly addresses how differences in variability, correlation strength, and bias reflect the roles of temporal sampling and spatial representativeness, and how these factors influence the relative performance of DS and SS observations in satellite validation. After revision, this figure is replaced by separate correlation analysis for DS and SS retrievals in Figure 13 and Figure 14.

Revised Text

Section 3.4 (Comparison of DS-SS Pandora and Satellite HCHO)

"A refined comparison between direct-sun (DS) and sky-scan (SS) retrieval geometries (Figs. 13–14) indicates that their relative performance depends strongly on the statistical metric considered. DS retrievals consistently exhibit higher correlation with TROPOMI, particularly at urban-influenced sites (e.g. Singapore: DS $r \approx 0.50$–0.52 vs SS $r \approx 0.43$–0.44; Bangkok: DS r up to ≈ 0.51 vs SS r up to ≈ 0.32), reflecting their stronger sensitivity to short-term variability and localized variability. However, SS retrievals can simultaneously achieve lower error magnitudes, as demonstrated at Bangkok where SS exhibits reduced RMSE and MAE compared to DS despite slightly lower correlation. Quantitatively, this improvement is substantial, with RMSE reduced by ~10–30 % and MAE by ~5–20 % in SS relative to DS depending on the collocation window, indicating a more consistent agreement in absolute column magnitude. This apparent inconsistency arises from differences in spatial representativeness: DS measurements sample a narrow atmospheric column and therefore capture fine-scale variability that enhances correlation but increases mismatch with the spatially averaged satellite pixel, whereas SS retrievals integrate multiple viewing directions and better approximate the satellite footprint, leading to reduced RMSE and MAE. This behaviour is most pronounced in urban environments with strong spatial gradients, while in low-HCHO regions such as Agam and Pontianak, SS retrievals show comparable or slightly improved agreement across both correlation and error metrics (e.g. Pontianak: SS $r \approx 0.40$ vs DS $r \approx 0.39$, with RMSE $\sim 3.4 \times 10^{15}$ vs $\sim 5.1 \times 10^{15}$ molecules cm^{-2}). Overall, these results demonstrate that DS retrievals are not universally superior; rather, DS and SS provide complementary strengths, with DS better capturing temporal variability and SS offering improved spatial representativeness for satellite validation in heterogeneous tropical environments."

Comment 3.32

Section 4.1: This discussion and figure 9 should still be under results. You are presenting new information.

Response 3.32

We thank the reviewer for this suggestion. We agree that the distinction between results and discussion should be clearly maintained. In the revised manuscript, the content associated with this analysis has been reorganized to ensure that the presentation of results and their interpretation are appropriately separated. The quantitative results are now presented within the Results section, while the Discussion section focuses on the interpretation and broader implications of these findings. After revision, this figure is now replaced by Figure 9 and Figure 12, which show the bias as a function of Pandora short-timescale variability.

Revised Text

Section 3.4 (Bias-variability Relationship)

"The bias–variability relationships (Figs. 9 and 12) further highlight fundamental differences in retrieval behaviour. For OMI, correlations between Pandora sub-daily variability and absolute bias are generally weak or inconsistent (e.g. $r = -0.07$ to 0.08 at Bangkok, $r = -0.34$ to -0.19 at Agam), indicating that OMI errors are not strongly linked to local temporal heterogeneity. In contrast, TROPOMI exhibits clearer and more physically consistent relationships, particularly at Singapore (DS: $r = 0.33$; SS: $r = 0.63$) and Pontianak (DS: $r = 0.35$; SS: $r = 0.27$), where increased short-timescale variability leads to larger satellite–ground discrepancies. Moreover, TROPOMI maintains lower overall error magnitudes compared to OMI, with RMSE typically below ~9×10^{15} molecules cm^{-2} and MAE below ~5×10^{15} molecules cm^{-2} at most sites. These results indicate that while TROPOMI remains sensitive to sub-pixel variability, its errors are more physically interpretable and systematically linked to atmospheric heterogeneity, whereas OMI discrepancies are dominated by coarse spatial resolution and representativeness limitations."

Comment 3.33

Figure 10: I don't think a and b are necessary. the colors and markers on this figure are difficult to read. Instead of SZA versus HCHO column, try SZA versus uncertainty.

Response 3.33

We thank the reviewer for this helpful suggestion. We agree that the original SZA-based analysis and figure presentation were not optimal for addressing the main objectives of the study. In the revised manuscript, the SZA-dependent analysis has been removed and de-emphasized, and the associated figure has been replaced by diagnostics that more directly reflect retrieval uncertainty and representativeness effects. As a result, Figure 10 is now removed and the issues related to panel labeling, marker readability, and the choice of SZA-based visualization no longer arise.

Comment 3.34

General/Conclusions:

Why not include GEMS in this analysis? The time range should be perfect and there are several figures showing the hourly Pandora columns.

Response 3.34

We thank the reviewer for this important comment and fully agree on the relevance of GEMS for this study, particularly given its high-temporal-resolution observations over Southeast Asia. In the revised manuscript, GEMS has been explicitly included in the analysis, alongside OMI and TROPOMI. The comparison framework has been expanded to incorporate GEMS in the time-series analysis, correlation assessment, and variability diagnostics. The inclusion of GEMS allows us to better evaluate the role of temporal sampling and diurnal variability, which is particularly important in tropical environments.

Revised Text

Section 3.5 Role of High-Temporal-Resolution Observations: Insights from GEMS HCHO Retrievals

Comment 3.35

Further development into understanding the data quality (and the data products) is necessary. For example, it is not made clear in the manuscript that DS is total column, while sky-scan is only the lower portion of the troposphere. The Pandoras are also pointing in different directions throughout the day in the DS mode.

Response 3.35

We thank the reviewer for this important comment and agree that clearer description of the Pandora retrieval characteristics is necessary. In the revised manuscript, we have clarified that direct-sun (DS) retrievals represent total column HCHO along the solar beam, while sky-scan (SS) retrievals represent a tropospheric column derived from multi-angle scattered radiation measurements, with sensitivity dependent on retrieval configuration and atmospheric conditions. We have also added clarification that DS observations are obtained by tracking the Sun throughout the day, resulting in changing viewing geometry and sampling direction, which can introduce additional variability in the observed column due to spatial heterogeneity. In contrast, SS observations provide a more spatially integrated measurement over a broader field of view.

Revised Text

Section 2.1 (Pandora Data Description)

"Direct-sun (DS) retrievals provide total column HCHO along the solar beam, whereas sky-scan (SS) retrievals represent a tropospheric column derived from multi-angle scattered radiation measurements, with sensitivity that depends on retrieval configuration and atmospheric conditions."

Section 4 (Discussion – strengthened interpretation)

"The application of uncertainty-based quality control improves the robustness of Pandora observations, while the separation of DS and SS retrievals reveals complementary strengths in capturing variability and spatially representative column structure. The observed differences between DS and SS retrievals reflect the interplay between measurement geometry and atmospheric heterogeneity, with DS capturing localized variability and SS providing a more spatially integrated representation of the atmospheric column."

Comment 3.36

Several figures could be removed/combined.

Response 3.36

In the revised manuscript, the figures have been carefully reviewed and reorganized. Redundant figures have been removed or consolidated. The main text now focuses on the key figures that directly support the primary conclusions.

Comment 3.37

This paper is presented as a comparison between Pandora and OMI, however the methods used for the comparison require further justification.

Response 3.37

We thank the reviewer for this important comment and agree that the comparison methodology required clearer justification in the original manuscript. In the revised manuscript, the comparison framework has been fully redesigned and explicitly justified. The analysis is now based on:

(i) an uncertainty-based quality control protocol (Rawat et al., 2025),

(ii) a rigorous DS–SS intercomparison using temporally matched pairs (instead of hourly or daily averages), and

(iii) a physically consistent collocation approach based on nearest-time matching and overpass-centered averaging.

In addition, the study has been expanded to include TROPOMI and GEMS, providing a multi-sensor context that strengthens the interpretation of satellite–ground differences. The revised methodology is now clearly described and consistently applied throughout the manuscript.