

Review on “Direct-sun versus Sky-Scan Pandora Formaldehyde Retrievals: Implications for OMI Validation in Tropical Southeast Asia”

General comment:

This manuscript investigates the differences between Direct-sun (DS) and Sky-scan (SS) Pandora HCHO retrievals and their implications for OMI validation over tropical Southeast Asia. The topic is relevant for satellite validation over the study domain; however, in its current form the manuscript lacks a clear motivation and several methodological choices are not sufficiently justified. The comparison between DS and SS observations is not rigorously addressed, and the satellite–Pandora comparison raises concerns regarding timing consistency, spatial representativeness, and robustness of the statistics. While the use of OMI is justified by its long-term data record (while only one year of OMI data is used in the analysis), many of the Pandora observations analyzed fall within the operational period of newer Geo-satellite instruments. Therefore, the use of GEMS, which was a key motivation for establishing the PAN-Asia Pandora network, would be highly recommended and more appropriate for the study region. In addition, several figures appear to contain unrealistic values, suggesting insufficient data screening and quality control. As a result, key conclusions presented in the abstract, such as the claim that Sky-scan retrievals systematically perform better than Direct-sun observations and that DS HCHO columns are strongly influenced by episodic enhancements, are not convincingly supported by the analysis. Substantial revisions are therefore required to clarify the methodology, improve data screening, and provide robust evidence for the conclusions.

Comments

Line 78-79: The direct-sun (DS) retrieval assumes negligible scattering, resulting in a nearly uniform sensitivity to the HCHO column regardless of its vertical distribution. Therefore, the statement that DS measurements have higher sensitivity to near-surface pollution is not accurate. Additionally, the cited reference does not support this claim. The authors should revise this statement and provide an appropriate explanation of the DS sensitivity characteristics. The citation to Herman et al. (2009) is also not appropriate in this context, as sky-scan Pandora observations were not available at that time.

Line 81–82: The statement “Despite these fundamental differences, most previous validation studies have implicitly treated Pandora HCHO as a single product” is unclear. Direct-sun (DS) and sky-scan (SS) Pandora HCHO retrievals have different sensitivities and are generally treated separately in validation studies. However, Rawat et al. (2025) proposed an approach to combine DS and SS observations into a single product by accounting for column biases and differences in integration time. The authors should clarify this statement and distinguish between studies that treat DS and SS separately and approaches that explicitly combine the two datasets.

I am surprised that the manuscript mentions several satellite missions such as TROPOMI, TEMPO, and Sentinel-4 but does not discuss GEMS, which was a key motivation for deploying Pandora instruments in the Asian domain under the PAN-Asia network. In addition, it is unclear why the analysis focuses only on OMI data when more recent satellite products such as TROPOMI and GEMS are available. The authors should justify the use of OMI alone or consider incorporating these newer datasets, which provide improved spatial and temporal coverage for validation studies.

It is difficult to use Figures 2 and 3 to intercompare the direct-sun (DS) and sky-scan (SS) observations. A more appropriate approach would be to compare temporally (within 5-10 minutes) matched DS and SS measurements (e.g., nearest observations) in a scatter plot to better assess the consistency and performance of the two observing modes.

For Pandora data quality: it appears that the authors use all flagged data (0, 1, 2, 10, 11, 12). However, medium- and low-quality data can sometimes contain large uncertainties and may require additional filtering (e.g., based on uncertainties, fitting WRMS or other quality criteria). At the same time, strictly removing all data flagged as 12 can sometimes eliminate a large portion of the dataset if only the highest-quality PGN flags are retained. Therefore, applying additional quality screening, similar to the approach proposed in Rawat et al. (2025), would likely strengthen the robustness of the analysis rather than using all flagged data without further evaluation. Additionally, the Pandora quality flag has three broad groups (Assured, Not-Assured, and Unusable). However, Pandora quality flags contain more detailed information, specifically, the units digit (0, 1, 2) indicates high-, medium-, and low-quality retrievals, respectively, while the tens digit indicates the Not-Assured. Thus, data flagged as 0 or 10 are generally considered high-quality and suitable for scientific use, whereas 1 or 11 and 2 or 12 indicate medium and low quality and require additional scrutiny for use (Gebetsberger et al., 2022).

For Figure 4, the time-series analysis is presented using multiple subplots, which makes it difficult to clearly see the overall temporal behavior. I recommend consolidating the information into a single figure showing the hourly and daily variations, and additionally including monthly averages to better illustrate the temporal patterns in the dataset.

Table 4: The temporal averaging for Pandora is described as 00:00–23:00 LT, which is unclear since Pandora instruments only operate during daytime under sunlight conditions (unless moon). Currently, there are no nighttime observations. Please clarify.

Although I understand the motivation for averaging OMI pixels to reduce noise and increase the number of collocated observations, the motivation for averaging all Pandora measurements over the entire day is not clearly justified. Pandora observations exhibit strong diurnal variability, and averaging over the full day may mask important temporal variability relevant for satellite validation. Recent work has emphasized the importance of carefully accounting for spatial

representativeness when using Pandora data for satellite comparisons. Also Park et al., (2026) shows that increasing the collocation radius generally improves the R^2 between Pandora and TROPOMI for HCHO, whereas the opposite behavior is often observed for NO₂ due to its stronger spatial heterogeneity. I recommend that the authors provide a clear justification for this averaging approach. Alternatively, they could restrict the analysis to the afternoon Pandora observation windows (E3, E6, and E9). If robustness is a concern due to limited sampling, using daily averages may provide a more representative comparison. However, using morning averages to compare with OMI observations does not appear justified, given the differences in overpass time and the strong diurnal variability in trace gas columns.

For the comparison between Pandora and OMI in Figure 5, the reported best performance for E2 and E8 appears questionable in terms of both temporal representativeness and robustness. It is unclear why morning Pandora observations would provide the best agreement with OMI, given the differences in overpass timing and the strong diurnal variability of trace gas columns. In addition, deriving statistical relationships from very limited numbers of collocated data points may not provide robust conclusions. I recommend that the authors consider using E6 or E9. If sampling robustness is a concern, E4 or E7 could also be considered, or the temporal matching window could be slightly relaxed to include additional observations. Similarly, the analysis presented in Figure 7, which relies on a single satellite pixel, may not be sufficient. A spatial averaging approach using multiple nearby pixels would likely provide a more representative comparison. Finally, the analysis shown in Figure 8 also appears to suffer from both robustness issues and potential timing mismatches, which should be carefully reconsidered to ensure meaningful satellite–Pandora comparisons.

It is also unclear why the OMI analysis is limited to only one year (2024). This choice is not justified in the manuscript, particularly since several Pandora sites have longer periods of data availability.

Again, it is difficult to understand why the SZA analysis is relevant for the OMI comparison, since OMI observations occur near early afternoon when the Sun is generally high in the sky, even during winter in tropics. It is also unclear whether the SZA values used in the analysis are derived from OMI or from Pandora observations. If the SZA values are based on Pandora measurements, the authors should clearly explain how the daily averaged Pandora SZA is used to justify the SZA dependence in the OMI comparison.

Section 4.3, titled “*Bias Correction and Retrieval Optimization*,” appears misleading, as no clear evidence of bias correction or retrieval optimization is presented. The satellite comparison results shown in the previous section contain substantial errors, and no robust or consistent bias statistics are demonstrated to support the claims made in the section.

References:

Park, J.-U., Lim, S., Hanisco, T. F., Abuhassan, N., Place, B. K., Pandey, A., Cede, A., Tiefengraber, M., Gebetsberger, M., Park, J., Choi, J., Crawford, J. H., Song, C.-K., & Kim, S.-W. (2026). Global analysis of nitrogen dioxide and formaldehyde column densities from the Pandora global network: Variability and implications for satellite validation. *Remote Sensing of Environment*, 335, 115249. <https://doi.org/10.1016/j.rse.2026.115249>

Rawat, P., Crawford, J. H., Travis, K. R., Judd, L. M., Demetillo, M. A. G., Valin, L. C., Szykman, J. J., Whitehill, A., Baumann, E., and Hanisco, T. F.: Maximizing the scientific application of Pandora column observations of HCHO and NO₂, *Atmos. Meas. Tech.*, 18, 2899–2917, <https://doi.org/10.5194/amt-18-2899-2025>, 2025.

Gebetsberger, M., Tiefengraber, M., Cede, A.: Fiducial Reference Measurements for Air Quality, TN on Data Quality Flagging Generic Procedure Evolution, Pandonia Global Network, https://www.pandonia-global-network.org/wp-content/uploads/2023/01/LuftBlick_FRM4AQ_DataQualityFlagging_GenericProcedureEvolution_TN_2019008_v7.pdf (last access: 27 September 2023), 2022.