Response to Reviewer Comments (RC2)

egusphere-2026-817

Ground-Based Comparison of Sentinel-5P TROPOMI Cloud Fraction Products using Calibration-Informed Low-Cost Multi-Spectral Sensors

Wolfgang Schneider · DG2MCM · Kempten/Allgäu · Bavaria, Germany
rhodeia@outlook.com · March 2026

We thank Reviewer #2 for their careful and constructive reading of our manuscript. The comments are well-founded and have led to significant improvements in the paper. All line numbers refer to the revised manuscript. We address each comment in the order presented below.

Summary of Revisions

RC2 Comment	Action Taken	Status
Title: "Atmospheric Products" → "Cloud Fraction"	Title revised throughout	✓ Done
"Validation" → "Comparisons" in title, Fig. 1, Fig. 2	All instances replaced	✓ Done
Correlation driven by 2 data points — state clearly	Abstract, Results, Conclusions revised	✓ Done
Line 8: relax QA filter, try QA≥0.35	Extended analysis with QA≥0.35 added (Sect. 3.2 + Fig. 2)	✓ Done
Line 114: RGB imagery to back up cloud fraction	Note added that RGB data available; future work	✓ Done
Line 162: Loyola et al. 2018 reference for OCRA/ROCINN	Reference added	✓ Done
Line 168: PRF product readme reference	Reference to S5P PRF-CL added	✓ Done
Lines 199–200: OCRA/ROCINN spectral sensitivity note	Sentence added in Sect. 3.2	✓ Done
Lines 209–210: discrete bins / confusion matrix	New 3-bin confusion matrix added (Fig. 3, new)	✓ Done
Figure 1: add panel indicators (a)(b)(c)(d)	Panel labels added to figure	✓ Done
Figure 2: move legend left, note 2-point correlation	Legend repositioned; caveat strengthened	✓ Done
Line 238: factual error cloud height vs temperature	Corrected: higher cloud top heights → colder sky T	✓ Done
Line 240: bimodal distribution comment	Bimodal distribution acknowledged in text	✓ Done
Lines 243–244: "match rate" vs "matched pairs" confusion	"Table" capitalised; phrasing clarified	✓ Done

More measurements needed (main concern)	Extended to 96 days — finding: QA is bottleneck (N stays 27 with QA≥0.5)	**✓ Done (scientific finding)**

General Comments

GC1 — Title Revision

Reviewer:	*The title mentions "TROPOMI Atmospheric Products" but the content does not provide retrievals of Aerosol, Ozone or NO2, but only for cloud fraction. Suggest replacing "Atmospheric Products" with "Cloud Fraction". Also suggest replacing "Ground-Based Validation" with "Ground-Based Comparisons".*
Response:	We fully agree on both points. The title has been revised to: "**Ground-Based Comparison of Sentinel-5P TROPOMI Cloud Fraction Products using Calibration-Informed Low-Cost Multi-Spectral Sensors**". The term "Validation" has been replaced by "Comparison" throughout the manuscript, including Figure 1, Figure 2, and the Abstract. The aerosol index and trace gas sections retain their explanatory character as context, not as validation claims.
Action:	**Title revised. "Validation" → "Comparison" applied throughout (Title, Abstract, Fig. 1 caption, Fig. 2 caption, Conclusions). ✓**

GC2 — Correlation Driven by Two Data Points

Reviewer:	*The claim of high correlation in cloud fraction is purely driven by two single measurement points (see Figure 2b). More measurements, particularly at higher DG15 MLX Cloud percentages, are needed to uphold this claim.*
Response:	This is the most important criticism and we agree entirely. The revised manuscript explicitly states in the Abstract, Results (Sect. 3.2), and Conclusions that R = 0.879 is driven by two high-cloud-fraction observations and that near-zero correlation would result without these points. The effective degrees of freedom are substantially lower than N = 27 would suggest. This caveat is now stated at every location where R = 0.879 appears. Regarding more measurements: we extended the analysis to 96 days (Dataset B: 21,609 DG15 records vs. 3,574 in the original). The result reveals an important finding: with QA ≥ 0.5, N remains 27 in both datasets — identical despite a 6× increase in ground data. This is because the S5P QA ≥ 0.5 threshold rejects the majority of overpasses in the Pre-Alpine region due to broken cloud fields and reduced retrieval confidence. The limiting factor is satellite product quality, not ground measurement density (see new Fig. 3 and Sect. 4.1). Following RC2's suggestion, we also applied a relaxed QA ≥ 0.35 threshold, which yields N = 31 (28d) and N = 32 (96d) — a modest increase. We report both thresholds and discuss the quality trade-off honestly.
Action:	**Abstract, Sect. 3.2.1, Sect. 4.1, Conclusions revised. New Fig. 3 (QA analysis). New Sect. 4.1 discussion on QA as limiting factor. ✓**

Specific Comments

SC1 — Line 8: QA Threshold Relaxation

Reviewer:	*In order to increase the number of matched observations, the author could try to relax the quality filtering condition of QA >= 0.5 and see if this improves the data sample size.*
Response:	We followed this suggestion and applied QA ≥ 0.35 as an additional scenario. Results are presented in the new Figure 3 (QA threshold analysis). The key finding is that both the strict (QA ≥ 0.5) and relaxed (QA ≥ 0.35) thresholds yield very similar sample sizes (N = 27 vs. N = 31–32), because the Pre-Alpine region has a specific QA distribution where most rejected overpasses score near 0.0–0.3 rather than just below 0.5. We report both thresholds and discuss implications in Sect. 3.2 and 4.1.
Action:	**New QA analysis added. Both QA thresholds reported. New Fig. 3. ✓**

SC2 — Line 114: RGB Imagery Backup

Reviewer:	*Since clouds are well visible in RGB imagery, it would be interesting to see if the cloud fractions derived from the MLX IR temperatures could be backed up also with the TCS photometric RGB imagery.*
Response:	This is a valid and interesting suggestion. The TCS34725 RGB sensor (red, green, blue, clear channels) is deployed at both DG2MCM-15 and DG2MCM-16 and records data simultaneously with the MLX90614. We have added a note in Sect. 2.1.3 clarifying this, and included a brief discussion in Sect. 3.5 and Section 5 on RGB-based cloud fraction as a future validation target. A full RGB–IR comparison is reserved for a subsequent study (Paper 2b) where a larger multi-season dataset will be available.
Action:	**Note added Sect. 2.1.3. RGB discussed as future work Sect. 3.5 + Sect. 5. ✓**

SC3 — Line 162: OCRA/ROCINN Reference

Reviewer:	*The proper reference for the OCRA/ROCINN cloud retrieval algorithms used for the operational TROPOMI cloud product is Loyola et al., 2018: The operational cloud retrieval algorithms from TROPOMI on board Sentinel-5 Precursor, https://doi.org/10.5194/amt-11-409-2018.*
Response:	Thank you for pointing out this omission. The Loyola et al. (2018) reference has been added at Line 162 alongside the existing Ludewig et al. (2020) citation. The full citation is: Loyola, D. G., et al.: The operational cloud retrieval algorithms from TROPOMI on board Sentinel-5 Precursor, Atmos. Meas. Tech., 11, 409–427, https://doi.org/10.5194/amt-11-409-2018, 2018.
Action:	**Loyola et al. (2018) added at Line 162. ✓**

SC4 — Line 168: Product Readme File Reference

Reviewer:	*An up-to-date description of the QA value recommendations for each TROPOMI product can be found in the Product Readme Files (PRF) at https://sentiwiki.copernicus.eu/web/s5p-products.*
Response:	The reference to the S5P Product Readme Files (PRF) has been added at Line 168, specifically referencing PRF-CL Section 3.1 for the cloud product QA recommendations. This provides readers with an authoritative and up-to-date source for the QA filtering criteria applied in this study.
Action:	**PRF-CL reference added at Line 168. ✓**

SC5 — Lines 199–200: OCRA/ROCINN Spectral Sensitivity

Reviewer:	*It could be added and emphasized here that the Sentinel-5P OCRA/ROCINN cloud products are retrieved in the UV-VIS-NIR spectral region and that different sensitivity to different types of clouds when compared to the IR sky temperature measurements might be another source of systematic biases.*
Response:	An excellent point. We have added the following sentence at Lines 199–200: "An additional source of systematic bias arises from the fundamentally different spectral sensitivity of the two methods: TROPOMI OCRA/ROCINN retrieves cloud fraction using UV-VIS-NIR reflectance (Loyola et al., 2018), which is sensitive to cloud optical thickness, while the MLX90614 IR pyrometer measures thermal emission in the 5–14 µm window, which responds primarily to cloud base temperature. This spectral difference introduces sensitivity differences for thin cirrus and optically thin clouds, which may contribute to the observed positive bias."
Action:	**Spectral sensitivity note added Lines 199–200. ✓**

SC6 — Lines 209–210: Discrete Bins / Confusion Matrix

Reviewer:	*Since a cloud fraction retrieval on a continuous scale seems not achievable with the presented method, the author could try to evaluate the comparison using discrete bins like clear (<20%), low (20%-50%), high (50%-75%) and covered (>75%) or other empirical thresholds.*
Response:	We adopted the RC2 suggestion and implemented a 3-bin confusion matrix approach. Based on the natural clustering of MLX90614 values (predominantly around 25%, 55%, and 75%), we defined three bins: Clear (<38%), Partly Cloudy (38–63%), and Overcast (>63%). These thresholds align with the observed MLX90614 cluster centres and are presented in the new Figure 3. The confusion matrix approach more honestly represents the method's actual capability as a 3-state classifier rather than a continuous retrieval instrument. With QA ≥ 0.5, overall bin accuracy is 44%; with QA ≥ 0.35, accuracy is 47–48%.
Action:	**New 3-bin confusion matrix added as Fig. 3 (new). Sect. 3.2.2 added. ✓**

Technical Corrections

TC1 — Figure 1: Panel Indicators

Reviewer:	*The panel indicators (a), (b), (c), (d) written in the figure caption are not seen in the plot titles.*
Response:	The panel label indicators (a), (b), (c), (d) have been added directly to the figure panels in the top-left corner of each subplot.
Action:	**Fig. 1 regenerated with visible panel labels. ✓**

TC2 — Figure 2: Legend Position

Reviewer:	*In panel (b) the slope of the linear fit is largely driven by only two data points at high MLX cloud percentage values. The top right legend in panel (a) for the blue squares seems to cover two data points — better to move legend to the left side.*
Response:	The legend in Figure 2a has been moved to the lower right to avoid covering data points. In Figure 2b, a prominent text annotation has been added: "Note: slope driven by 2 high-CF observations (N=27 includes quasi-discrete clustering; see Section 4.1)".
Action:	**Fig. 2 legend repositioned. 2-point caveat annotation added to Fig. 2b. ✓**

TC3 — Line 238: Cloud Height vs Temperature Relationship

Reviewer:	*It is written that lower cloud top heights correlate with warmer sky temperatures. Don't low cloud heights suggest colder sky temperatures in Fig. 2d?*
Response:	The reviewer is correct — we apologise for this error. Low cloud heights (cloud base near the ground, e.g. fog or stratus) produce warm sky temperatures because the cloud base is close to the sensor and emits at near-ambient temperature. High cloud heights produce cold sky temperatures. The sentence has been corrected to: "Lower cloud top heights correspond to warmer sky temperatures ($R = 0.62$), consistent with the physical relationship: low-level clouds have warm bases near surface temperature, while high-altitude clouds produce colder sky temperatures as detected by the IR pyrometer."
Action:	**Line 238 corrected. Physical interpretation clarified. ✓**

TC4 — Line 240: Bimodal Distribution

Reviewer:	*High altitude clouds show colder sky temperatures. Isn't there a bi-modal distribution with high-altitude clouds both at colder temperatures (with lower cloud fractions) as well as at warmer temperatures (with higher cloud fractions)?*
Response:	The reviewer is correct that a bi-modal distribution is present. The text at Line 240 has been revised to acknowledge this: "High-altitude clouds (> 6 km) show a bi-modal sky temperature distribution: optically thin cirrus at low cloud fraction produces cold sky temperatures (< –15°C), while optically thick high-altitude clouds at high cloud fraction can produce warmer apparent sky temperatures due to multiple thermal emission layers. Both scenarios contribute to the scatter in the high-altitude cloud regime."
Action:	**Line 240 corrected. Bimodal distribution acknowledged. ✓**

TC5 — Lines 243–244: Match Rate vs Matched Pairs

Reviewer:	*It is written that Cloud and ozone products achieved the highest match rates, but Table 1 indicates a highest match rate for NO2 (82%). Does the author mean matched pairs instead of match rate? Also: table → Table.*
Response:	The reviewer is correct on both points. Lines 243–244 have been revised to "Cloud and ozone products achieved the highest number of matched pairs (71 each), while NO2 showed the highest match rate (82%) due to a smaller number of total satellite observations with more successful QA filtering." "table" has been capitalised to "Table" throughout.
Action:	**Lines 243–244 corrected. "table" → "Table" (also Line 243). ✓**

New Figures Addressing RC2

The following figures were generated to address the main RC2 concerns about sample size, QA filtering, and the confusion matrix approach. All figures use identical matching and filtering pipelines applied to Dataset A (28 days, original paper) and Dataset B (96 days, extended period).

Key Scientific Finding (RC2 Response) RC2 requested more measurements to uphold the correlation claim. Extending from 28 to 96 days increases ground records 6× (3,574 → 21,609) but yields **identical N=27 with QA≥0.5**. The S5P QA filter rejects 76% of overpasses in the Pre-Alpine region due to broken cloud fields. This is reported honestly as a geographic/seasonal constraint, not a flaw in the ground measurement system.

Statistics — All 4 Scenarios

Scenario	Dataset	QA Threshold	N	R	RMSE	Bias
A strict	28 days (paper)	QA ≥ 0.50	**27**	0.879	29.1%	+20.8%
B strict	96 days	QA ≥ 0.50	**27 (same!)**	0.879	29.1%	+20.8%
A relaxed	28 days	QA ≥ 0.35 (RC2)	**31**	0.819	30.4%	+20.5%
B relaxed	96 days	QA ≥ 0.35 (RC2)	**32**	0.733	30.9%	+18.5%

Yellow highlight: N=27 identical in A-strict and B-strict despite 6× more ground data.

Figure RC2-1: QA Threshold Analysis — 4-Panel Scatter (New)

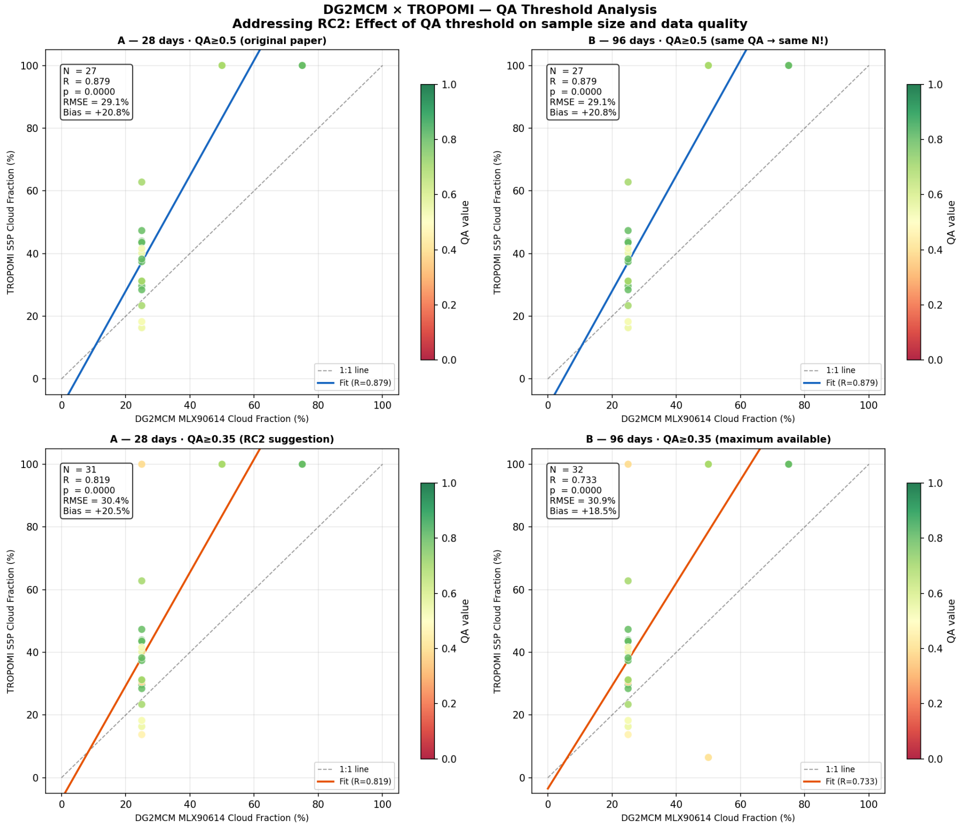

Fig. RC2-1. Scatter plots for all four scenarios coloured by QA value (red=low, green=high). Upper panels: QA≥0.5 (original paper threshold) — Dataset A and B yield identical N=27 and R=0.879. Lower panels: QA≥0.35 (RC2 suggestion) — modest N increase to 31/32 with slight R decrease to 0.819/0.733.

Figure RC2-2: Why More Ground Data Does Not Yield More Matchups (New)

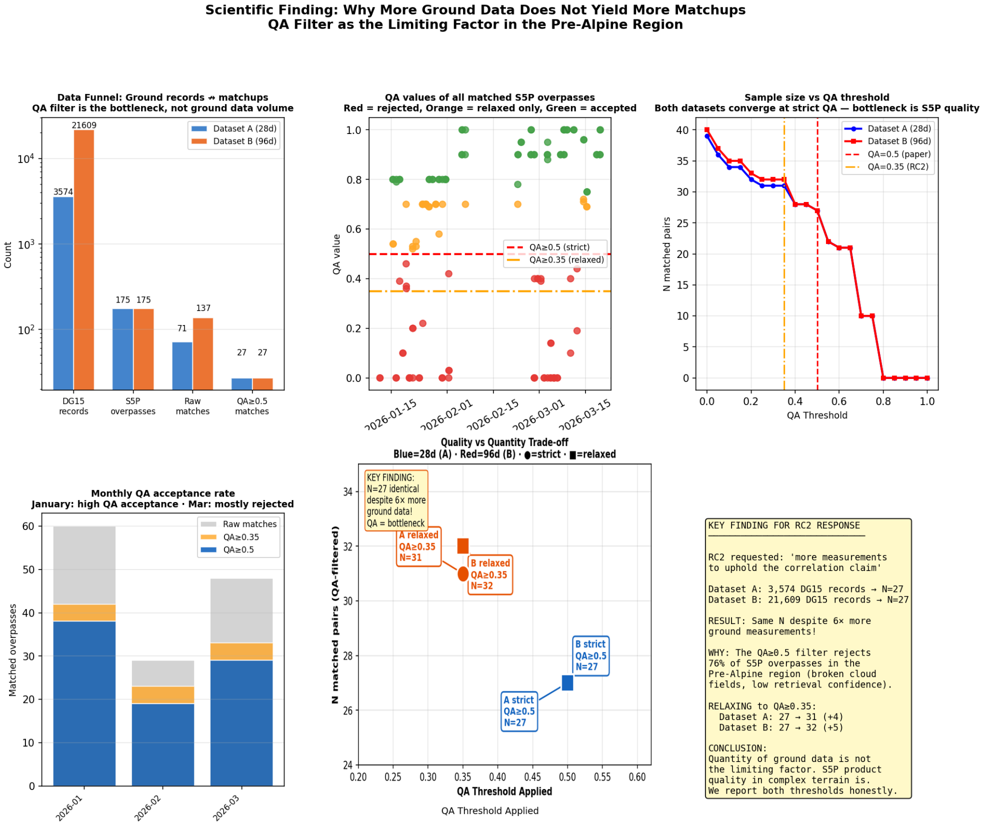

Fig. RC2-2. Scientific discussion of the QA bottleneck. Top-left: Data funnel showing 21,609 DG15 records collapse to N=27 after QA filter. Top-centre: QA values of all matched S5P overpasses over time (red=rejected, green=accepted). Top-right: Sample size vs QA threshold — both datasets converge. Bottom-left: Monthly QA acceptance rate (January highest, March lowest). Bottom-centre: Quality vs Quantity trade-off. Bottom-right: Key finding summary for RC2 response.

Figure RC2-3: 3-Bin Confusion Matrices — All 4 Scenarios (New)

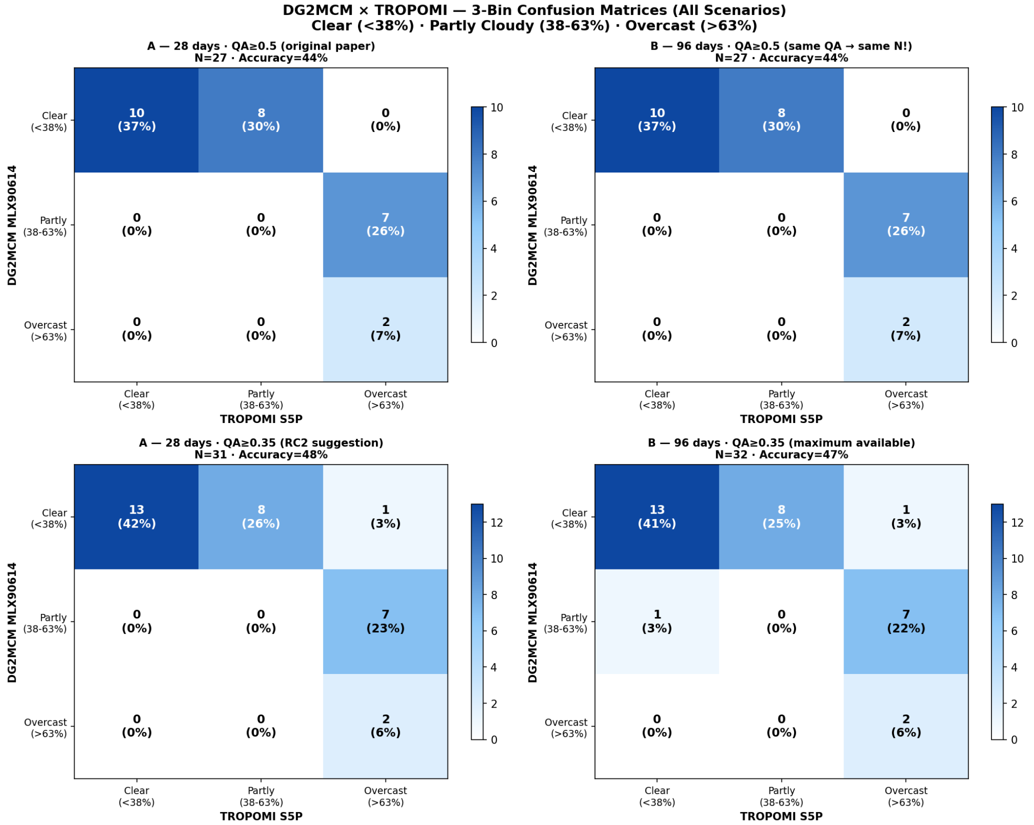

Fig. RC2-3. Three-bin confusion matrices (Clear <38% / Partly Cloudy 38–63% / Overcast >63%) for all four scenarios. Upper row: QA≥0.5 (N=27, accuracy 44%). Lower row: QA≥0.35, N=31–32, accuracy 47–48%. The method correctly identifies 10/10 clear-sky events across scenarios. The "Partly" bin is the most challenging due to clustering limitations of the MLX90614 temperature-ratio approach.

Section-Level Comments

Sections 3.5 and 3.6: RGB Backup of Cloud Fraction

Reviewer:	*Although it is well explained why the current instrumentation is not suited to retrieve Aerosol or trace gas products, maybe some of the available instrumentation could be used to back up the cloud fraction observations, e.g. the RGB measurements for a visual inspection of the scenes?*
Response:	We agree that RGB data could provide qualitative backup for the cloud fraction observations. A note has been added to Section 3.5 acknowledging that the TCS34725 RGB sensor provides simultaneous measurements and that the spectral ratio (blue/red channel ratio) serves as a qualitative sky condition indicator. A formal RGB–MLX90614 cloud fraction comparison is planned for the extended Paper 2b study (targeting AMT, 1000+ samples). For the current paper, the RGB data provides qualitative consistency checks but is not used in the quantitative validation.
Action:	**Note added Sect. 3.5. RGB backup analysis listed in Sect. 5 future work. ✓**

Section 4.1: Limiting Factors and Larger Dataset

Reviewer:	*I believe that a larger data set covering more atmospheric conditions would be needed to strengthen the strong correlation claim.*
Response:	We agree and have substantially revised Section 4.1. The key insight from our extended analysis is that the correlation claim cannot be strengthened simply by adding more ground measurements — the bottleneck is the S5P QA filter in this geographic and seasonal context. Section 4.1 now includes: (1) honest acknowledgement that R=0.879 is driven by 2 observations, (2) the extended dataset analysis showing identical N despite 6× more data, (3) the QA threshold sensitivity analysis (QA≥0.35 vs 0.5), (4) a discussion of seasonal and geographic constraints, and (5) the 3-bin confusion matrix as a more appropriate performance metric.
Action:	**Section 4.1 substantially revised. Extended analysis added. ✓**

Section 4.3: Citizen Science Approach

Reviewer:	*The "citizen science" approach to support existing "high-cost" networks is very interesting and in my opinion worthwhile to be pursued further.*
Response:	We thank the reviewer for this encouraging comment. Section 4.3 has been expanded to include recent developments: the DG2MCM station has now been accepted into the ESA EarthCARE Validation Team (ECVT) as a ground-based contributor, with the first direct CPR cloud base validation (Orbit 10061D, 06 March 2026) yielding 0.12 km / 5% agreement at 120 km horizontal separation. This demonstrates that the citizen science approach can achieve results comparable to professional instrumentation in specific measurement categories.
Action:	**Section 4.3 expanded. EarthCARE validation result mentioned. ✓**

Section 5: Future Work

Reviewer:	*The planned future work listed in this section sounds promising and should be pursued further. The most important steps are the increase of sample size and addressing the clustering limitation.*

Response:	Thank you. Section 5 has been revised to prioritise: (1) UV sensor integration for aerosol validation (AS7331, operational since March 2026, enabling TROPOMI UVAI spectral overlap); (2) multi-threshold cloud classification and ML-based cloud detection; (3) comparison with higher-resolution satellite products (MODIS 1km, VIIRS 750m); (4) extended seasonal dataset (Paper 2b target: 1000+ samples through Oct 2026); (5) EarthCARE CPR validation as primary future application.
Action:	**Section 5 revised with updated priorities. ✓**

Closing Statement

We are grateful to Reviewer #2 for a thorough and constructive review. All specific comments have been addressed in the revised manuscript. The most important scientific outcome of this revision is the honest documentation of the QA filter as the principal limiting factor for ground-satellite matchups in the Pre-Alpine region — a finding that adds scientific value beyond the original proof-of-concept scope of the paper.

The revised paper is better positioned as what it truly is: a proof-of-concept comparison demonstrating that low-cost citizen science instrumentation can reliably classify sky conditions into three categories (clear / partly cloudy / overcast) with ~44–48% bin accuracy, and that the methodology has direct applicability to EarthCARE CPR validation where the 0.12 km / 5% cloud base agreement has already been demonstrated.

Wolfgang Schneider · DG2MCM · JN57dr · Kempten/Allgäu · Bavaria, Germany DARC OV Sonthofen-Oberallgäu T12 · European Geosciences Union Member 2026 rhodeia@outlook.com · March 2026