Responses to Comments

We sincerely thank Editor Dr. Munchak, and two anonymous reviewers for the constructive and thoughtful comments.
Comments are in blue italic lettering, responses in black.

Reviewer 1 Comments

(1) The manuscript highlights the advantages of the hybrid BMCI-OEM framework over a BMCI-only approach. Could the authors comment on the computational performance of the hybrid retrieval framework? While BMCI is described as computationally efficient, the OEM component requires iterative optimization and numerical Jacobian calculations. It would be useful to know how often the OEM step is triggered in practice and what additional computational cost it introduces compared to a BMCI-only retrieval.

Thanks for the comments. The current OEM implementation is computationally intensive, which is primarily due to the numerical Jacobian calculation step. With the current version, each retrieval can take approximately 10-20 minutes, which is very slow. We are currently exploring alternative methods for faster Jacobian estimation and optimizing the code for the matrix calculations to improve computational efficiency. A discussion of this limitation has been added to the revised manuscript, as shown in Lines 554-556 in the track-changed document.

As for the likelihood of triggering the OEM step, the case study shown in Fig. 10 indicates that a large majority of the retrievals require the additional OEM step. This is partly because the selected testing cloud scenes contain dense clouds, which are generally sparsely represented in the retrieval database. In addition, the current threshold for triggering the OEM step may be too stringent. It would be interesting to see the actual likelihood of OEM activation when the algorithm is applied to global observations over entire orbits, such as the currently available AWS observations and future PolSIR/PMM observations. More comprehensive results on this aspect will be reported in future work.

(2) Please consider including the spatial resolution/footprint size of the various instruments used in the study, as well as the expected characteristics of the future missions (e.g., C²OMODO footprints <5 km have been reported in the literature). In addition, please provide the radar sensitivity (minimum detectable reflectivity, dBZ), as this information is essential for understanding the sensitivity of the reference retrievals.

Information on the footprint size of the C2OMODO and CoSSIR have been added, as shown in Lines 229-230 and in Lines 353-355 in the track-changed document. Also, the minimum detectable limits for each radar channel have been added, as shown in Lines 357-358.

(3) The manuscript demonstrates a notable sensitivity of the simulated brightness temperatures to the assumed ice particle habit. All simulations further assume randomly oriented particles and use only the first Stokes component. While polarization differences vanish at nadir, particle orientation can still modify the bulk scattering and extinction properties and therefore affect the simulated brightness temperatures. Given that the selected particle models are adopted from May et al. (2024), where

orientation effects are also discussed, could the authors briefly comment on the potential implications of neglecting particle orientation for the retrieved ice microphysical properties? In addition, errors associated with neglecting particle orientation and polarization effects may become increasingly important at larger viewing angles, which are relevant for the planned cross-track satellite observations, particularly for future dual-polarization conically scanning radiometers.

Thanks for the comments. Particle alignment and polarization observations are indeed important remaining issues that have not been addressed in this study. As you mentioned, even at the nadir angle where polarization effects vanish, horizontally aligned ice particles can produce larger extinction and therefore lower the TB. At off-nadir angles, horizontally aligned ice particles increase the extinction at H and decrease the extinction at V, resulting in the generation of Stokes Q, which cannot be reproduced under the random orientation assumption. The effect of particle orientation is important even for the first Stokes I only. For randomly oriented particles, the ice cloud is effectively homogeneous in terms of Stokes I, which means the extinction coefficient remains nearly identical from nadir to large off-nadir viewing angles. For horizontally aligned particles, however, the extinction coefficient for Stokes I decreases with increasing viewing angle, indicating that the cloud becomes more transparent at off-nadir angles than at nadir. This results in an interesting phenomenon that we observed in our simulations for the high-frequency channels: Stokes I may increase as the viewing angle changes from nadir to off-nadir, even though the propagation path through the cloud becomes longer.

We are still evaluating how to incorporate polarization into the retrieval framework developed in this study. Ideally, a scattering database for horizontally aligned particles would be used to perform Stokes I/Q simulations. However, the computational burden may make this approach less feasible. We will report our progress in future studies. The limitation of the random-orientation assumption and the lack of addressing polarization have been added to the revised manuscript, as shown in Lines 550-554 in the track-changed document.

(4) The current implementation assumes pencil-beam radiative transfer simulations and evaluates only nadir observations. Please provide information on the horizontal resolution of the simulations and discuss the extent to which the reported performance may be affected by the neglect of finite instrument footprints and off-nadir viewing geometries. Given that both PolSIR and PMM-C2OMODO are cross-track scanning radiometers, beam-filling effects may become increasingly important at submillimeter frequencies.

In this respect, please delete the word "slightly" from the following statement: "All results presented here assume nadir-only observations: we acknowledge that both the PolSIR and PMM-C2OMODO radiometers are cross-track scanners, and results at off-nadir angles can be slightly different" as differences are to be expected.

Thank you for the comments. The current study adopts a pencil-beam simplification, and the effects of footprint size and horizontal resolution have not been addressed. The current database was generated by directly inputting the vertical profiles of clouds and atmospheric properties from CloudSat products into a 1D radiative transfer model and simulating the TB at nadir. As you mentioned, beam-filling effects become increasingly important with increasing viewing angle, and the pencil-beam assumption becomes less reasonable for off-nadir observations. We are currently updating the a priori database by both expanding its sample size and incorporating the effects of footprint size and horizontal variability. The

limitation of pencil-beam assumption has been added, as shown in Lines 547-550 in the trach-changed document. Also, the word "slightly" in the sentence regarding the retrieval performance at off-nadir angles has been deleted, as shown in Line 244.

Suggestions for future work: The paper reports the largest residual brightness temperature discrepancies for the channels closest to 325 GHz and for the most strongly scattering cloud scenes. For future developments of the retrieval framework, the authors may wish to explore alternative particle model selections. For example, McEvoy et al. (2026, EGUsphere preprint) adopted the same particle habit framework as May et al. (2024), but introduced an additional ICON-hail particle model after identifying discrepancies between simulated and observed submillimeter brightness temperatures in strongly scattering deep convective conditions. In this context, analysing the joint behaviour of the 183 and 325 GHz channels (e.g., through scatterplot diagnostics) may provide a useful tool for evaluating and refining the choice of particle models used in the retrieval framework.

Thanks for the suggestions. The latest knowledge about particle habit selection is much appreciated. As mentioned above, we are currently updating the retrieval database, and the suggested new particle habit will be included in the updated database.

Reviewer 2 Comments

#1. (Lines 46-50) This sentence mixes missions, satellite platforms, and instruments. Please revise the sentence for clarity.

The sentence has been revised to differentiate the instrument and the missions, as shown in Lines 46-52 in the track-changed document.
(Note: All line numbers below refer to the track-changed version of the manuscript.)

#2. (Line 55) Brath et al. (2018) uses an ensemble of standard neural networks and may not be an example of a quantile regression neural network. Could the authors please check whether this reference is appropriate here?

The sentence has been revised to differentiate between NN and QRNN, and the reference has been updated, as shown in Lines 55-60 in the track-changed document.

#3. (Line 124) Because the previous sentence describes the synergistic CloudSat-CALIPSO 2C-ICE product, it would be helpful to clarify that 2B-CWC-RO is a radar-only product.

The sentence has been revised to include the information that the 2B-CWC-RO is a radar-only product, as shown in Lines 127-128.

#4. (Lines 138-139) Prigent et al. (2017) describes TESSEM2. Also, could the authors clarify whether TELSEM2 was used for land emissivity? Aires et al. (2011) describes the TELSEM, which does not cover frequencies above 200 GHz.

Thanks for pointing this out. The simulation uses the built-in TESSEM2 model for ocean surfaces and the TELSEM2 model for land surfaces in ARTS. According to the ARTS documentation server, TELSEM2 is applicable over the frequency range from 5 to 900 GHz, and Aires et al. (2011) is cited as the reference.

The TELSEM2 documentation is available at the following link: https://atmtools.github.io/arts-docs-2.4/docserver/methods/surfaceTelsem.html

The sentence has been revised to update the ocean and land surface models to TESSEM2 and TELSEM2, respectively, as shown in Lines 142-143, but the reference remains unchanged.

#5. (Line 159) The right-hand side of Eq. (3) gives the variance. Please revise the text or equation accordingly.

Eq. (3) has been updated to use standard deviation instead of variance.

#6. (Line 168) I understand the point the authors are making, but inflating the NEDT does not change the actual measurement accuracy. It instead weakens the observational constraint in the retrieval. The term "measurement accuracy" may therefore be misleading. Please consider revising this wording.

The sentence has been revised to 'weakens the observational constraint imposed on the retrieval' following your suggestion, as shown in Lines 172-173.

#7. (Lines 180-181) It is unclear what is meant by "mapped onto the CDF of a standard normal distribution." Is the inverse standard normal CDF applied to the rank values?

The inverse standard normal CDF is a better expression in this context. The sentence has been revised accordingly, as shown in Lines 185-186. Thanks for your suggestion.

#8. (Eq. 5) The first K in Eq. (5) should be transposed.

Eq.5 has been revised to use the transpose of K.

#9. (Lines 227-228) The term "cases" refers to profiles?

The term 'cases' has been changed to 'profiles' for clarity, as shown in Line 234.

#10. (Lines 244-245) The range from 0.5*IWP_true to 2*IWP_true corresponds to -50% to +100%

The ±100% has been revised to -50% to +100% in both the text and the figure, as indicated in Line 251 and Figure 1.

#11. (Line 265) Please revise ", with most values falling below 10^1 and are …." for grammatical consistency.

The word 'are' has been deleted for grammatical correctness, as shown in Line 271.

#12. (Eq. 6) Eq. (6) seems to be incorrect. The first K should be transposed. If I understand correctly, K represents the sensitivity of the simulated TBs to the state vector. When K=0, the observations provide no information about the state, so the averaging kernel should also be zero. However, Eq. (6) gives the identity matrix in this case. This equation may need revision.

Thanks for pointing this out. The equation was typed incorrectly and has now been revised. I've also checked the code and confirmed that the formula used in the code is correct.

#13. (Line 309) It seems that Sa and Sp are the covariance matrices, but the text describes them as PDFs.

The sentence has been revised to specify the covariance matrices, as shown in Line 315.

#14. (Line 355) CloudSat was launched in 2006, so data from winter 2005 would not be available.

The time range has been revised to 2006-2019, as indicated in Line 364.

#15. (Lines 395-402) The interpretation based on excess ice mass looks reasonable, but I'm curious why the BMCI-only mismatch increases toward the centers of the water vapor absorption lines. Some explanation of this spectral pattern would be useful.

This is an interesting point, especially considering that the channels farther away from the absorption frequency are generally more sensitive to ice clouds. My interpretation is that the channels closer to the absorption line are more sensitive to ice clouds near the cloud top. This may explain why the BMCI retrieval shows many artificial ice clouds near the cloud top that should not exist, as shown in Figure 8. The OEM step improves the simulation accuracy of the channels closer to the absorption line, which leads to a significant reduction in these artificial ice clouds near the cloud top.

A sentence has been added to provide this explanation, as indicated in Lines 407-409.

#16. (Line 406 and Figure 7) Reflectivity differences should be expressed in dB.

The unit of the reflectivity differences has been corrected to dB in both the text and the figure, as indicated in Line 417 and Figure 7.

#17. (Lines 459-460) Since Xy^2 is dimensionless, I'm not sure what is meant by "fall below or close to the measurement uncertainties."

The χ_y^2 follows a chi-squared distribution with M degrees of freedom, with a mean of M and a standard deviation of $\sqrt{2M}$, where M is the number of channels. As the number of CoSSIR channels used here is 8, we therefore expect the χ_y^2 statistic to follow a chi-squared distribution with a mean of 8 and a standard deviation of 4. The sentence has been revised to express the points more clearly, as shown in Lines 471-472.

#1. (Line 16): HIWARP -> HIWRAP

Corrected, as shown in Line 16.

#2. (Line 28): ice clouds properties -> ice cloud properties

Corrected, as shown in Line 28.

#3. (Line 41): IceCube was launched in 2017, not 2016

Corrected, as shown in Line 41.

#4. (Line 65): samplings -> samples

Corrected, as shown in Line 68.

#5. (Line 221): The numbers of the 183.31 GHz and 325.15 GHz channels are reversed.

Corrected, as shown in Line 226.

#6. (Line 294): from -> between

Corrected, as shown in Line 300.

#7. (Line 295): corresponds -> correspond

Corrected, as shown in Line 301.

#8. (Line 316): lowers -> decreases

Corrected, as shown in Line 322.

#9. (Line 320): contributes -> contribute

Corrected, as shown in Line 327.

#10. (Line 345): four subchannels near 325 GHz -> three subchannels near 325 GHz

Corrected, as shown in Line 352.

#11. (Figure 11 caption): vertical -> vertically

Corrected, as shown in the Figure 11 caption.

#12. (line 515): HIWARP -> HIWRAP

Corrected, as shown in Line 527.