Manuscript number: amt-2021-414
Full title: Cloud optical properties retrieval and associated uncertainties using multi-angular and multi-spectral measurements of the airborne radiometer OSIRIS
Author(s): Matar et al.

The paper investigates the retrieval uncertainties of liquid cloud properties, specifically cloud optical thickness (COT) and droplet effective radius (Reff), associated with various error sources using an optimal estimation (OE)-based retrieval procedure. The retrieval error covariance matrix is decomposed into the one originating from individual sources of errors, which makes it possible to evaluate the quantitative retrieval uncertainty estimation for each error source, including measurement error, model parameter errors, and error due to the assumptions in the forward model. The authors apply this framework to multi-angular and multi-spectral measurements made by OSIRIS airborne radiometer with a focus on liquid clouds. The results indicate the forward model assumptions that do not consider in-cloud heterogeneous profiles and 3D radiative transfer effects induce the largest uncertainties in the retrievals, followed by the measurement errors. While the model parameter errors provide the least impact on the retrieval uncertainties that are less than 0.5% of the retrieval quantities. These error estimates and retrieval procedures will be useful for the further 3MI observations.
While this paper is well written in the introduction and methodology part, I am concerned about the representativeness of the results and the analysis flow of this work. Please find my comments below. The topic presented in this paper is suitable for Atmospheric Measurement Techniques. I recommend Major Revisions to reconsider the manuscript for publication.

Major comments

Representativeness

First of all, the paper discusses only the relative standard deviations (RSD) of the retrieval variables. Why don't the authors discuss the retrieval *biases* associated with the measurement errors, model parameters, and assumptions in the forward models? In particular, the vertical heterogeneity and 3D radiative effects would induce substantial biases in the retrievals. The different magnitude of errors among observational bands and angular directions would cause impacts on the cloud property retrievals (through changing the sensitivity weights). However, the relative standard deviations of the retrieval variables do not tell you how much the retrieval variables are biased. If the authors want to solely focus on the retrieval RSD for each error source, then it can be more appropriately achieved through numerical experiments based on synthetic simulations of observational signals based on this framework or through incorporating the root-mean-square error (RMSE) that includes the retrieval bias information and can be theoretically derived from the bias and RSD. Alternatively, the authors may consider to additionally perform cloud property retrievals based on error covariance matrix that considers ALL sources of errors (i.e., measurement errors, model parameter error, and forward model errors) to see how different the cloud property retrievals are compared to measurement error-only cases.

In addition, the uncertainties and fixed parameters of the source of errors seem to be determined in different ways. The cloud top height value and its uncertainty are derived based on collocated lidar observations (i.e., representing the CALIOSIRIS campaign), the effective variance seems to be arbitrarily determined, and the wind speed value and uncertainty may represent global statistics. This gives me a question on the representativeness of this study. If the authors make the results useful to the future 3MI mission, the uncertainty (specifically, the standard deviation of a model parameter) of each error source (in particular, the model parameter errors) should represent those of global climatology. Specifically, I feel that the cloud top height and effective variance uncertainties may be a bit too small, so the retrieval RSD might be underestimated (although the general conclusion the little contribution of model parameter uncertainties to retrieval RSD may not change).

For these reasons, I suggest the authors clearly state what are the focuses of this paper in the last paragraph of the introduction and reconsider the experimental design of this work through, but not limited to, using numerical experiments, adding the retrieval biases to the current analysis, or other appropriate methods.

Analysis flow

While I appreciate the authors for describing the solid mathematical basis of the analysis procedure, I may have an argument on the use of COT and Reff retrievals obtained based only on measurement noise in the error covariance matrix for the following analysis. This can overfit the retrieval variables to obtain an optimal solution (i.e., $J \sim n_y$) because radiative signal *perturbations* (here what I mean is any of radiative perturbation induced by atmospheric-cloud properties naturally occur in the real world) originating from other sources of errors are partly explained with the *biased* retrieval values. If such perturbations are significant (and yes, it is significant particularly for vertical heterogeneity and 3D effects), then the sensitivity of the other error sources to the retrieval quantities are obtained from the sensitivities to the retrieval variables at the biased cloud property conditions. Ideally, numerically generated cloud property fields (such as Large Eddy Simulations) would provide datasets for the evaluations of the retrieval uncertainties based on the authors' flamework. As the true cloud properties are not available in the observed cloud field, it is not possible to address this issue based on the given observations. Therefore, at least, the authors should discuss a potential bias of the retrieval uncertainty evaluations based on this framework.

Minor comments

1. Title: Actually, the authors use only two bands for the cloud property retrievals, and therefore, "bispectral" measurements would be more descriptive rather than "multi-spectral" measurements?

2. Line 241, Eq. (7): Why do the authors use a linear scale of the cloud optical thickness for a state vector element, not a logarithmic scale? Although the relation between solar reflectance and COT is linear at very small COT conditions ($\tau \ll 1$), it is in general quite non-linear over most of the COT range, which makes the retrieval process slow and may degrade the convergence, and also may limit the representativeness of this results to optically thin cases.

3. Section 3.3: I do not find the values of *a priori* error covariance matrix in the manuscript. Please briefly mention what values are chosen.

4. Line 338 "$\sigma_{\text{alt}} = 0.16$" :: This represents only the cloud properties observed during the CALIOSIRIS campaign.

5. Line 333 "$\sigma_{\text{veff}} = 0.003$" :: How did you get this value? It seems too small. Please site references that support this quantity.

6. Lines 372–376: In addition, the collision-coalescence process can provide a larger droplet radius at the lower part of clouds, which are observed from CloudSat. As vertical heterogeneity is important to the cloud property retrieval, the authors may consider a better representation of cloud profiles using a better cloud profile parameterization (e.g., Saito et al., 2019).

7. Lines 416–420: As the authors assume the flat cloud top, which reduces some of cloud 3D effects such as illuminating and shadowing effects, I had an impression that the authors may focus on lateral photon transport effect here. If so, it would be better to rephrase 3D with lateral photon transport or state "3D" regarded as the lateral photon transport.

8. Figure 4: The readers cannot recognize if there are optimal/non-optimal solutions from these plots. If it is non-optimal, a set of cloud retrievals may not adequately explain the measured signals. I suggest the authors add the cost function distributions in addition to these two plots.

9. Lines 483–484 *"3D effects due to solar illumination do not appear in the retrieved cloud properties"* :: This is an obvious statement as the cloud top is assumed to be flat, which removes the shadowing and illuminating effects. Please state that this error evaluation focuses solely on the lateral photon transport effect.

10. Lines 539–540: The cost function divided by the number of measurement signals (J/n_y) is a comparable quantity among mono-angular- and multi-angular-based retrievals.

11. Lines 559–560" Why are these uncertainties reduced for multi-angular cases?

12. Line 610: If the authors state that *"the uncertainties related to the measurement errors is implementable in an operational algorithm,"* the uncertainty evaluations should be based on global climatology of cloud, surface, and atmospheric properties. A limited case (i.e., a granule of a cloud scene) may not be adequate to state so.

13. Typo and grammatical errors: Please proofread the main body of the manuscript again. I have found several grammatical errors, e.g., Lines 557, 571, and possibly more.

Reference

Saito, M., P. Yang, Y. Hu, X. Liu, N. Loeb, W. L. Smith Jr., and P. Minnis, (2019) An efficient method for microphysical property retrievals in vertically inhomogeneous marine water clouds using MODIS–CloudSat measurements, *J. Geophys. Res. Atmos.*, 124, 2174-2193.