Review of the article intitled: “Above Cloud Aerosol Detection and Retrieval from Multi-Angular Polarimetric Satellite Measurements in a Neural Network Ensemble Approach” by Yuan et al. for publication in AMT.

Anonymous reviewer.

Opinion on the publication:

Major review: Significant corrections, clarifications, and additions are required before this article can be accepted.

General description:

This article describes an algorithm for detecting aerosols above clouds using the spectral, angular, and polarized radiance data provided by the POLDER spaceborne instrument. The method is potentially applicable to future other instruments that will provide similar data. They use some of the concept and work previously conducted on aerosol detection above clouds with the POLDER instrument using a LUT-type method, but this time with an AI method, using neural networks.

The article details a three-step methodology for remote sensing of aerosols in cloudy scenes: liquid cloud detection, aerosol property retrieval above clouds, and a posteriori radiance simulation. Details of the datasets employed in training the neural network, specifically regarding aerosol and cloud particle properties, are provided. The method is evaluated using synthetic data, then applied to a year of POLDER instrument data. Aerosol properties are compared to those from an established clear-sky algorithm, though direct comparison is limited as the latter covers the total atmospheric column, not just above clouds. A brief qualitative discussion compares the new method's results with those from POLDER's operational algorithm for aerosols above clouds.

General review:

The use of AI methods in aerosol remote sensing, especially for complex cloudy scenes, is a promising and justified approach. While the geophysical results are quite convincing, I have reservations regarding several aspects of the article:

-The description of the neural network's characteristics and the assessment of its performance is too brief.

-The state of the art requires significant enrichment and a thorough refresh.

-The algorithm's results have not been quantitatively validated or compared against equivalent products.

More specifically, the individual networks aren't evaluated; only the overall solution is. Key methodological choices are not justified, and some information on the network structure, are missing, making replication difficult. Correlations between retrieved and test parameters are often quite low, and the paper lacks sufficient metrics in tables for readers to adequately assess the method's accuracy. Furthermore, the AI references are outdated, which is surprising given the rapid advancements in this field.

It's also important to include comprehensive references to existing operational methods for detecting aerosols above clouds, particularly those developed for passive and active instruments. The community has made significant comparisons between these active and passive methods, both for case studies and global applications, and the associated articles should be thoroughly cited. For the POLDER/PARASOL method (Waquet et al., 2013; Peers et al., 2015), a review of its technical aspects and geophysical results is needed. Technical details of this method would better explain the differences between the results presented in this article and the operational POLDER algorithm.

Finally, is seems that the proposed method doesn't account for sub-pixel cloud heterogeneity effects. This could lead to biases, potentially overestimating aerosol optical depth above clouds. At the very last, the limitations of using plane-parallel radiative transfer code for aerosol remote sensing in cloudy scenes should be mentioned.

My detailed comments and questions follow below.

Abstract.

Please add the wavelength(s) used for ACAOT, AE and SSA

Introduction.

You compare your results with those from the method previously developed for POLDER (Waquet et al., 2013, Peers et al., 2015). The parameters retrieved with your method and many of the underlying concepts, such as using MODIS data to identify and remove cirrus clouds and restricting retrievals to high cloud fraction and large COT values, are similar to those of the previously cited method.

So, this section requires improved referencing and description of the operational aerosol above cloud algorithm previously developed for POLDER/PARASOL, including the associated available product and validation efforts. The AERO-AC product, with its DOI, is globally available for 5 years of POLDER data, which is worth noting for the reader. https://www.icare.univ-lille.fr/aero-ac/

Some of the following explanations should be incorporated into the manuscript (see also my additional comments at the end of this review).

- Initially, the Waquet et al. (2009, 2013) method determined above-cloud aerosol optical thickness and Ångström exponent exclusively from polarization measurements. This was achieved using a look-up table (LUT) approach combined with a decision tree strategy.
-The method was then improved by including additional total radiance measurements (Peers et al., 2015) to simultaneously retrieve the above cloud aerosol single scattering albedo and the cloud optical thickness of the below cloud layer (COT).
-The associated global product is referred to as AERO-AC (Waquet et al., 2020)
- The aerosol above cloud properties are only retrieved in case of homogeneous optically thick (COT > 3) and liquid water clouds. Cloud fractional covers and cloud edges are removed. Cirrus above liquid water clouds are also filtered and different quality criteria are eventually applied to improve the products.

Please add the following references:

Peers, F., Waquet, F., Cornet, C., Dubuisson, P., Ducos, F., Goloub, P., Szczap, F., Tanré, D., and Thieuleux, F.: Absorption of aerosols above clouds from POLDER/PARASOL measurements and estimation of their direct radiative effect, Atmos. Chem. Phys., 15, 4179–4196, https://doi.org/10.5194/acp-15-4179-2015, 2015.

Waquet F., Peers F., Ducos F., Thieuleux F., Deaconu L., A. Chauvigné and Riedi, J.: Aerosols above clouds products from POLDER/PARASOL satellite observations (AERO-AC products), doi:10.25326/82, 2020.

Please mention the methods that use active measurements to retrieve aerosol properties above clouds. Different methods (standard methods and advanced methods like the "depolarization ratio method") were developed for CALIOP and various products are available (see Jethva et al., (2014) and Deaconu et al. (2017))

It's also important to highlight the research community's dedication to validating and intercomparing their passive and active aerosol-above-cloud products**.** This has involved rigorous work, ranging from in-depth case study analyses (Jethva et al., 2014) — supported by airborne sun-photometer data (Chauvigné et al., 2021) — to comprehensive global scale analyses (Deaconu etal., 2017).

Please add the following references:

Jethva, H., O. Torres, F. Waquet, D. Chand, and Y. Hu (2014), How do A-train sensorsintercompare in the retrieval of above-cloud aerosol optical depth? A case study-based assessment, Geophys. Res. Lett., 41, 186–192, doi:10.1002/2013GL058405.

Deaconu, L. T., Waquet, F., Josset, D., Ferlay, N., Peers, F., Thieuleux, F., Ducos, F., Pascal, N., Tanré, D., Pelon, J., and Goloub, P.: Consistency of aerosols above clouds characterization from A-Train active and passive measurements, Atmos. Meas. Tech., 10, 3499–3523, https://doi.org/10.5194/amt-10-3499-2017, 2017.

Chauvigné, A., Waquet, F., Auriol, F., Blarel, L., Delegove, C., Dubovik, O., Flamant, C., Gaetani, M., Goloub, P., Loisil, R., Mallet, M., Nicolas, J.-M., Parol, F., Peers, F., Torres, B., and Formenti, P.: Aerosol above-cloud direct radiative effect and properties in the Namibian region during the AErosol, RadiatiOn, and CLOuds in southern Africa (AEROCLO-sA) field campaign – Multi-Viewing, Multi-Channel, Multi-Polarization (3MI) airborne simulator and sun photometer measurements, Atmos. Chem. Phys., 21, 8233–8253, https://doi.org/10.5194/acp-21-8233-2021, 2021.

Line 57: "*Section 5 shows the data processing of one year (2008) PARASOL measurements and comparison with adjacent PARASOL-RemoTAP clear-sky aerosol retrievals*."

A comparison with a similar algorithm would have been more relevant, given the inherent differences between aerosol concentrations integrated over the total atmospheric column (including low-altitude aerosols like marine aerosols) and those corresponding to aerosols above clouds.

Suggestion: The comparison between clear-sky and above-cloud aerosol retrievals could also focused on the fine mode Aerosol Optical Thickness (AOT). Such a comparison seems more relevant especially for biomass burning particles, which are predominantly fine mode and often found in elevated layers as for instance over the Southeast Atlantic region.

2. Data Description / section 2.3

Line 77: *"Here in this work, a pixel is marked as liquid phase only when the fraction of liquid-cloud-flagged 1-km-resolution MODIS pixels within a 6km × 6km PARASOL grid cell is larger than 80%."*

In Waquet et al. (2013), cloud optical thickness standard deviation was derived from 1-km-resolution MODIS retrievals within PARASOL pixels. They applied criteria to select only homogeneous POLDER pixels, based on spatial variability in cloud properties.

This allows to reduce the plan parallel effects that impact the modeling of polarize radiance especially in the cloud bow region (Cornet et al., 2013). This effect may result in false detection of aerosol above clouds (positive bias in the ACAOT)

Does your method control for sub-pixel cloud property heterogeneity by rejecting the most heterogeneous pixels? or is this neglected? Please clarify this point.

Please add Cornet et al., 2013 in the list of reference.

Cornet, Celine & C.-Labonnote, Laurent & Szczap, F. & Deaconu, Lucia-Timea & Waquet, Fabien & Parol, Frederic & Vanbauce, Claudine & Thieuleux, François & Riedi, J.. (2017). Cloud heterogeneity effects on cloud and aerosol above cloud properties retrieved from simulated total and polarized reflectances. Atmospheric Measurement Techniques Discussions. 1-25. 10.5194/amt-2017-413.

At the very last, mention the inherent limitations of using plane-parallel radiative transfer code for aerosol remote sensing in cloudy scenes

Line 110: "Only the measurements with a minimum of 14 angles are considered for the NN training, in order to evade from a variable-sized input vector to the NN or, as an alternative, an input vector with missing data."

This sentence is not unclear to me. Could you rephrase it or provide more explanation?

Section 3.2: Neural network training.

Line 159: *"To increase numerical efficiency and reduce memory usage during the training process, we choose the "neural network ensemble" approach (Hansen and Salamon, 1990)"*

Why did you choose the neural network ensemble? It typically requires significant data, computational power, and memory, which appears to contradict your goal of *"increasing numerical efficiency and reducing memory."*

Also, the reference Hansen and Salamon (1990) is quite old. Are there any more recent references on neural network ensembles?

How do you justify the use of an ensemble approach compared to using a classical method?

Please correctly write out the three proposed architectures:

- Show diagrams of the architectures.
- Present the hyperparameters for each architecture.

Describe the dataset for each step: what is used as input, the validation/test split, and include a table summarizing this information.

Also, it would be interesting to see the training curves for both validation and learning, so we can see the performance of your NNs

Line 169: *"The Adam optimizer (Kingma and Ba, 2014) is used to minimize the mean root square error (RMSE) loss function."*

Could you please specify the settings used for the Adam optimizer?

Line 123: *"In the training set, 20% of the samples represent the situation where the aerosol layer is located above the cloud top, in order to improve NN's ability to produce liquid and ice cloud fractions in areas of interest for this study. A pixel will be further processed"*

8 million data points, of which only 20% met the conditions. Why not use the correct number of data points directly if you're going to reduce it afterwards?

Line 121: *"with more cloud fractions close to 1 in order to acquire better sensitivity at almost fully cloudy cases"*

"Does this limit your reliable retrievals to areas with 100% cloud coverage? If so, please mention it. It would be useful to summarize the limitation(s) of your method in the conclusion section and abstract.

You mention that your state vector includes the cloud top altitude. Is this actually retrieved with your method? Have you compared your cloud top height retrievals with concomitant CALIOP data? If so, what is the robustness of your retrieval? What are the assumed aerosol base and top altitudes in your RT code?

Line 115: *"The first NN (liquid cloud mask) takes intensity, degree of linear polarization (DoLP), and viewing geometries (SZA, VZA, RAA and scattering angle) as input and outputs liquid cloud fraction and ice cloud fraction separately"*

The name of your first neural network, "liquid cloud mask," is a bit confusing. Since you're using it to estimate both liquid cloud fraction and cirrus cloud fraction, it seems to do more than a simple liquid cloud mask.

Also, how is your mask performing?

Line 143: *"The intensity and DoLP, as a function of wavelength and viewing angle, are compressed using a principal component analysis (PCA) before the training. A total of 25 principal components are retained for radiance and 33 for DoLP."*

Is the use of PCA indispensable? Please justify its inclusion, as its benefit is not immediately apparent when combined with a deep neural network.

Line 156: *"It should be noted that the NN forward model is not a complete forward model. It only works for pixels fully covered by a liquid cloud without any radiative contribution from the surface and is designed only for the purpose of goodness-of-fit assessment for above cloud aerosol retrievals."*

I'm not convinced the third network is truly necessary. Is it sufficiently accurate for predicting both total radiances and polarized radiances? How is its performance evaluated? It might be discarding valid retrievals if this NN is not accurate enough.

Line 161: *"The final output is the average of the outputs from all the ensembles"*

For the second NN, what are the discrepancies between the 16 networks? Are these discrepancies significant?

Line 169: *"and batch training with a batch size of 12,000"*

As the first reviewer noted, this value seems unusually high compared to what's reported in the literature. Please clarify.

Section 4: synthetics measurements

Figure 2 lack sufficient detail to evaluate the method's performance. Could you provide more metrics?

For instance, can you add linear fit results on the curves in Figures 2? and the number of considered points? it will be helpful.

For the results shown in Figure 2: Both absolute and relative Mean Absolute Errors (MAEs) should be provided. The results should be presented in tables.

Figure 2-e and Figure 2-h show the results with synthetic retrievals for the Ängström Exponent (AE).
I am surprised to see that the AE is systematically low biased for fine mode aerosols and high biased for coarse dust aerosol and the correlation coefficients are very low (<0.3).

I would expect to see random results scattered around the one-to-one line, similar to the general test results shown in Figure 2b.

Does this imply that your architecture is not adequately dimensioned to retrieve AE for extreme size distributions (e.g., purely fine or coarse modes)? If so, should the training be enhanced for these extreme scenarios? Such extreme conditions are particularly representative of satellite observations for aerosols located above clouds.

Line 185: *"For AE and SSA, an additional mask of retrieved ACAOT > 0.2 is applied."*
-Please specify the wavelength for the ACAOT considered here.

Line 194: *"The retrievals are always masked by a retrieved liquid cloud fraction larger than 0.8"*
Could you recall the spatial resolution of your cloud mask?

Line 195: Same comment, please add wavelength for the ACAOT

Line 202: *"Over ocean, we see an opposite effect (except for very small COT), because the contribution from the ocean is relatively small and a smaller COT would even enhance the relative contribution of the aerosol signal compared to the cloud signal."*

Did you account for the surface wind speed and sun-glint in your method?

5.1 Comparison between PARASOL-NN above cloud aerosol retrievals and adjacent RemoTAP clear-sky aerosol retrievals

Similar to Figure 2, Figure 4 would benefit from additional metrics to properly evaluate the comparison results.

As previously discussed, the RemoTAP clear-sky algorithm results are not directly comparable with the above cloud aerosol properties retrieved with the present. It would have been more interesting to compare with existing aerosol above clouds available products.

Line 207: *"the data are aggregated at the same 1◦ × 1◦ grid cell".*

Could you also provide a comparison between clear-sky and above-clouds retrievals for a case study (e.g., a daily product for a portion of an orbit)? This is also important to show the spatial variability in the retrieved aerosol above clouds properties obtained with your method.

For Figure 5, please adjust the color scale for the ACAOT. It's currently difficult to discern differences for ACAOT values between 0 and 0.1 (most of the values …). A histogram of ACAOT would be also very useful.

In Figure 5: What is the wavelength for the ACAOT?

Line 233: There seems to be an error in the article citation.

Please cite the paper by Waquet et al. (2013b) that presents a geophysical analysis of the global aerosol properties above clouds using POLDER by season for 2008. This study is directly comparable to yours (see Figure 1 in Waquet et al., 2013b).

Waquet, F., F. Peers, F. Ducos, P. Goloub, S. Platnick, J. Riedi, D. Tanré, and F. Thieuleux (2013b), Global analysis of aerosol properties above clouds, Geophys. Res. Lett., 40, 5809–5814, doi:10.1002/2013GL057482.

To avoid confusion, please differentiate between the two Waquet et al., 2013 (a) (remote sensing method) and (b) (geophysical analysis) references

Waquet, F., Cornet, C., Deuzé, J.-L., Dubovik, O., Ducos, F., Goloub, P., Herman, M., Lapyonok, T., Labonnote, L. C., Riedi, J., Tanré, D., Thieuleux, F., and Vanbauce, C.: Retrieval of aerosol microphysical and optical properties above liquid clouds from POLDER/PARASOL polarization measurements, Atmospheric Measurement Techniques, 6, 991–1016, https://doi.org/10.5194/amt-6-991-2013, 2013a

From Lines 236 to 241: The comparison of your results with those of Waquet et al. (2013) is too succinct and qualitative. I would favor a more quantitative comparison, at least for some case studies.

Line 245: *"We have to remark that our AE in regions between 45◦− 60◦N and 45◦− 60◦S is∼ 0.8, which differs largely from∼ 1.8 in Waquet et al. (2013), despite the good agreement of our above cloud AE with the adjacent clear-sky AE in these latitudes."*

This funding is interesting and deserves more investigation.

Please add this information in the manuscript: the above-clouds AOTs associated with an AE of 1.8 in Waquet et al. (2013a) method for the 45°−60°N region are typically low (<0.05 at 865 nm), and even lower for the 45°−60°S region (<0.03 at 865 nm)

My opinion is that the ACAOTs are probably too low for effective aerosol type identification.

-What are your ACAOT values for these cases (i.e., cases with an AE of about 0.8)? Please add the corresponding ACAOT map to Figure 6

- Line 245: our AE in regions between 45◦− 60◦N and 45◦− 60◦S is∼ 0.8

What would be the source of these particles located above clouds?

For such retrieved AE values (AE of about 0.8), this means that your algorithm retrieves a mixture of non-spherical mineral dust and fine mode particles.

Is your clear-sky algorithm also detect non-spherical coarse mode (mineral dust) over these regions for adjacent cases?

What would be the source of these mineral dust particles located above clouds over the 45◦− 60◦S region in the south hemisphere?

Additional comments:

It would be useful to remind the reader of the operational retrieval strategy previously developed for POLDER. This will help them understand any potential differences between that method and your current approach. Some of the following explanations should be incorporated into the manuscript.

Initial Test Phase: Initially, the algorithm focuses on fine mode aerosols (considering six distinct models) and utilizes only forward and side scattering angles (scattering angle less than 130°). This range is chosen because, for Cloud Optical Thickness (COT) greater than 3, the polarized radiance reflected by the cloud is low (or weakly negative) and stable, becoming independent of cloud microphysics and optical thickness. Therefore, any additional polarization observed within this scattering angle range indicates the presence of a lofted aerosol layer above the clouds. (For a schematic view, refer to Figure 4 in Waquet et al., 2013a).

Full Retrieval Phase: If the retrieved above-cloud Aerosol Optical Thickness (AOT) from this initial test exceeds 0.1 at 865 nm, the algorithm proceeds to search for the best-fitting model among *all* available models. This comprehensive set includes the six fine modes plus a bimodal non-spherical mineral dust particle model, and the retrieval uses all available data across all scattering angles.

Conversely, if the above-cloud AOT is less than or equal to 0.1 at 865 nm, the AERO-AC algorithm solely retrieves the fine mode AOT and fine mode Angstrom Exponent (AE) using only observations for scattering angle smaller than 130°. So, for these situations, the AE values are the constrained between 1.6 and 2.4. This previous restriction on using the cloud bow and dust model in the AERO-AC algorithm was implemented to prevent false detection of AC events above clouds, particularly when the above-cloud atmosphere appears pristine based on initial tests.

Regional Limitations: It's important to note that in the regions between 45°–60°N and 45°–60°S, the above-cloud AOTs retrieved by the AERO-AC algorithm are likely too low to facilitate accurate aerosol type identification. This is a probable explanation for the observed differences between the mean AERO-AC results and those derived using the present method.