Response to reviewer 2

We would like to thank the reviewer for his/her important comments and suggestions.

Abstract.

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

The wavelengths are added in the revised manuscript.

Introduction.

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.

☐ 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.

In the revised manuscript, we added the suggested references and the related explanations to the introduction (line 40) and data description (line 85)

☐ 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.

The references are included (line 42 and 85).

- ☐ 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))

Those methods are included in the paper (line 45).

- ☐ 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.

The references above are included and the work from the community is highlighted in the revised manuscript.(line 52 and 232)

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.

We have added a subsection (5.2, to the paper with a comparison with the AERO-AC product to the paper. We believe both comparisons (i.e. to nearby clear-sky retrievals and AERO-AC) retrievals have their own specific relevance. For example, we expect that ACAOT is generally correlated with total AOT, but ACAOT should be smaller. For situations with larger AOD (>0.2), the intrinsic aerosol properties are expected to have many similarities between above-cloud cases and clear-sky cases, and plausible explanations can be found for remaining differences (e.g. at high AE we expect the above-cloud AE to be slightly larger than total column AE, because it is less influenced by Sea Salt).

We have also. added the fine mode AOT comparison between clear-sky and above-cloud aerosol retrievals (Figure 6 in the revised manuscript).

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.

In our method, there is no such control, but we expect that the goodness-of-fit criterion filters out many of these situations, because they will cause variations between viewing angles that cannot be modeled by the 1D forward model (Stap et al., 2015; 2016). We added a discussion on this topic to the revised manuscript and included the reference to Cornet et al. (section 3.2, line 126).

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

Now the limitations are mentioned in the paper.(section 3.2, line 126)

- ☐ 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?

PARASOL-POLDER can observe a ground pixel at up to 16 angles, but the number of viewing angles varies over the different L1C pixels. The majority of pixels observers a ground pixel at 14 angles and that is what we trained our NN for. To train an NN for a variable size of the input vector is very challenging. In principle a better approach would be to train separate NNs for different sizes of the input. We clarified that in section 2.1, line 70.

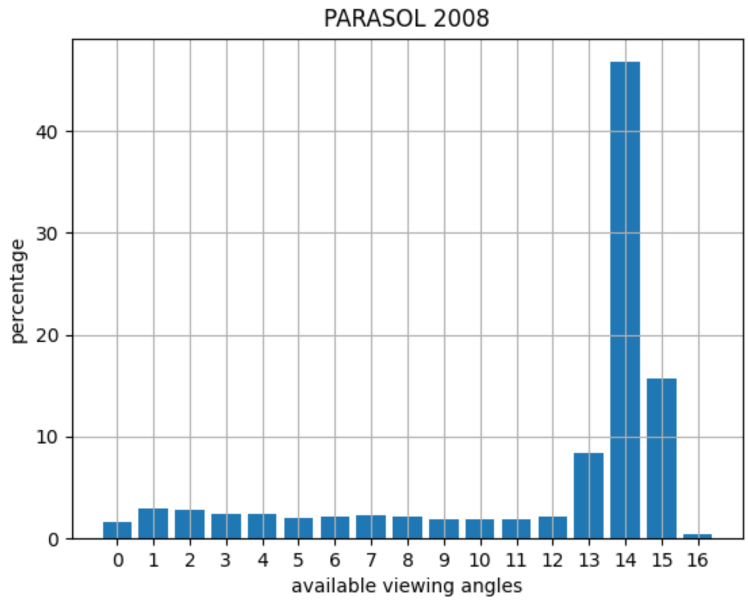

Fig 1. Histogram of PARASOL available viewing angles per pixel.

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.”

Using “neural network ensemble” approach can significantly reduce the NN’s overfitting and increase NN’s generalization (Ortega et al, 2021). Based on our experiments (both in this paper and previous studies), a similar performance can be achieved by training all samples in one go or separating them into several ensembles (e.g., an NN trained with 16 million samples or 16 NNs trained with 1 million samples each). However, the latter (NN ensemble method) requires far less total training time and memory, and different ensembles can be trained simultaneously on different computing nodes individually. Based on the mentioned merits, we chose NN ensemble approach. It is true that when applying the NN, an ensemble approach has a higher computational cost, but still this is negligible compared to full physics algorithms.

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

M.A. Ganaie, et al, (2022) wrote a review of the development of neural network ensemble strategies, including badging, boosting, stacking, etc (mainly on classification application). In our paper, we use the approach in Hansen and Salamon (1990) (therein they use the majority voting scheme for classification, while we use the averaging strategy). We add Ganaie, et al (2022) as reference as well in the revised manuscript (line 179).

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

Please see my answer above about the justification for the ensemble approach.

☐ 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.

Diagram (Fig 1 in SI, NN ensemble structure) and table (Table 1 in SI, three NNs' details, e.g, input, output, etc) are added to show the NN architectures as well as the inputs and outputs of the different NNs in the SI of the revised manuscript.

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

Here we show the loss function of training set and holdout set from one ensemble in the ACA retrieval NN, and it can be seen that the loss function on the two sets both converges well without overfitting features. We added these figures to the SI of the paper (Fig 2 in SI).

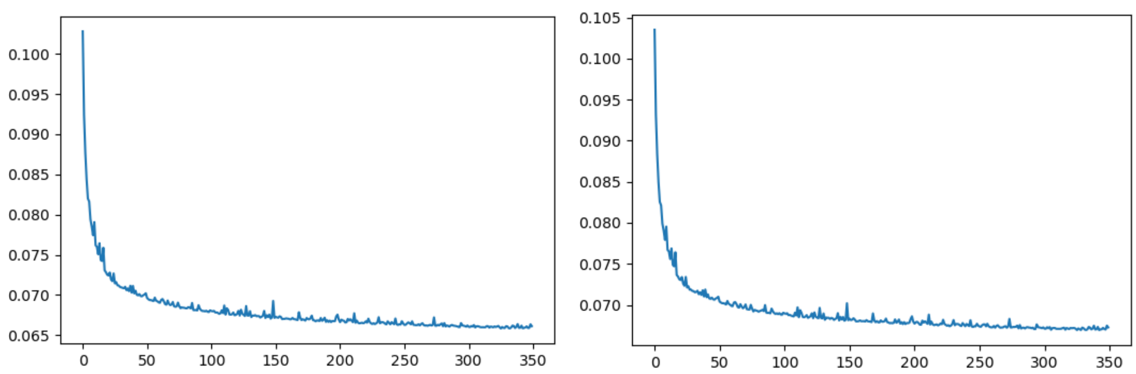

Fig 2. Loss function v.s training epoch on training set (left) and holdout set (right).

☐ 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?

The optimizer settings include: learning rate = 0.001, betas = (0.9, 0.999), eps = 1e-8 and weight_decay = 0. The settings are the default recommended settings. The adam optimizer is an adaptive optimizer and based on our experience, it is virtually not sensitive to the initial learning rate. We included this information in the revised manuscript (line 192).

- ☐ 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?
This line describes the training set of cloud mask NN. For the aerosol retrieval NN, all pixels in the training set are with aerosol above clouds. We have clarified this in the revised manuscript (line 160)

- ☐ 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.

This does limit reliable retrievals to large cloud fractions (CF > 0.80) but not just fully cloudy pixels. We clarified this in the conclusion of the paper (line 137).

- ☐ 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?

Yes, the ACA retrieval NN outputs the full state vector including cloud top height (CTH), but from the performance over holdout set (test set), the CTH is not well retrieved (correlation is 0.56 and RMSE is 600 (m)). Therefore, we didn't compare it with CALIOP data. The aerosol profile follows a Gaussian distribution with a fixed FWHM=2000 m, and we retrieve only the center altitude (aerosol layer height). We clarified it in the revised manuscript (line 103)

☐ 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?

We changed the name “liquid cloud mask NN” to “cloud mask NN”. The performance of the mask is shown in the figures below. The figures are also included in the SI of the revised article (Fig 3 of SI)

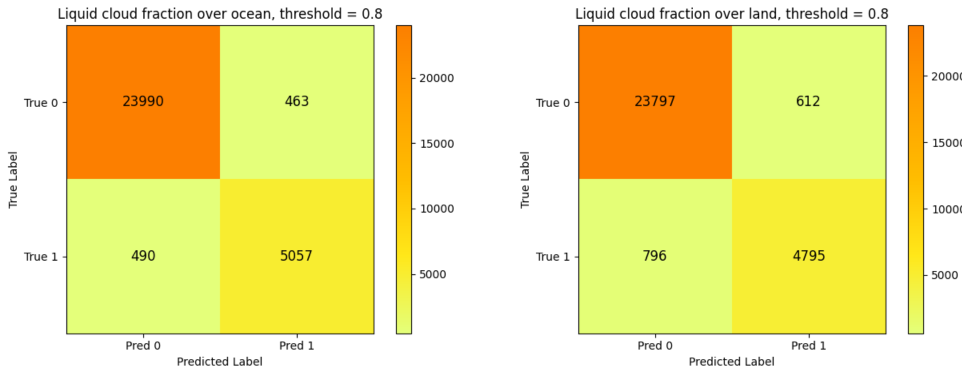

Fig 3. Confusion matrix of liquid cloud detection on the holdout set, “pred 1” means predicted liquid cloud fraction > 0.8, “true 1” means true liquid cloud fraction > 0.8.

☐ 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.

An acceptable result can be obtained without PCA, but using PCA makes the results slightly better (from synthetic test) as a way of denoising.

☐ 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.

Below we show the comparison of intensity and degree of linear polarization (DoLP) between NN forward model and RemoTAP forward model, at 565nm. The rstd (relative standard deviation) of intensity is 0.7% and the std (standard deviation) of DoLP is 0.0025, both of which are below the instrument measurement noise. This suggests that the NN forward model is good enough to replace the full physical model (RemoTAP) in estimation goodness-of-fit. We added these figures to the revised manuscript (Figure 2 of the revised manuscript).

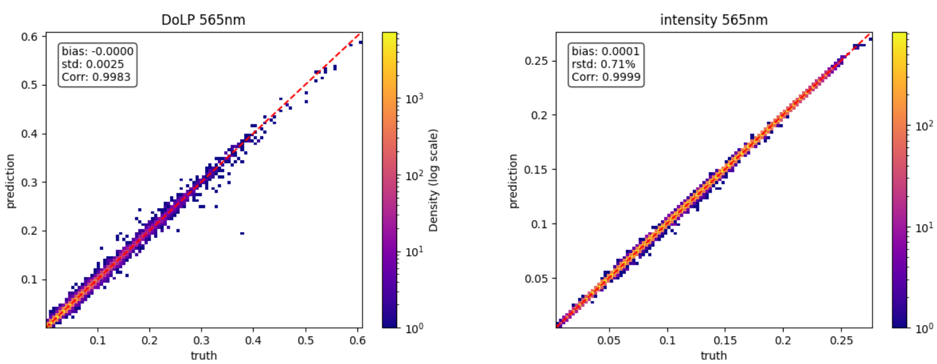

Fig 4. Intensity (left) and degree of linear polarization (DoLP, right) from NN forward model (prediction) and RemoTAP forward model (truth) at 565nm.

Additionally, the figures below show the comparison between RemoTAP clear sky retrieval and the NN ACA retrieval (as is in section 4) but without the goodness-of-fit chi2 mask derived from the 3rd network. It is clear to see the chi2 mask filtered out a lot of unphysical retrievals and improved the performance. The figures are included in the SI (Fig 4 of SI)

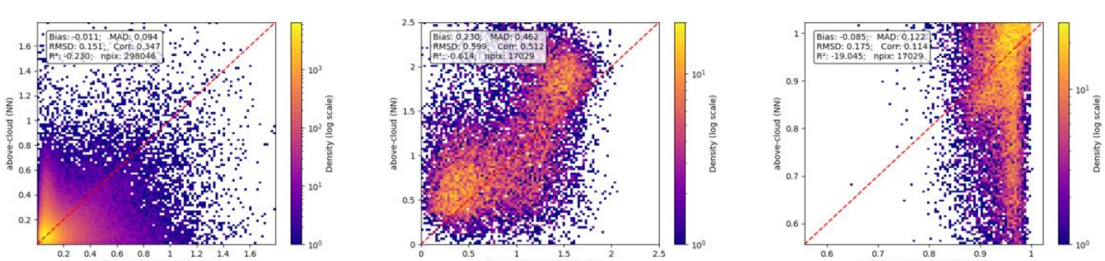

Fig 5. NN ACA retrievals v.s. adjacent PARASOL-RemoTAP clear sky retrievals. No goodness-of-fit mask applied. Other filters are the same as in section 4.1 of the paper.

☐ 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?

Below shows the ACAOT (550nm) mean retrieved value and the range across the different ensemble members from randomly chosen 100 pixels (1% of all) on the synthetic validation dataset used in the paper (both fine and dust mode aerosol). The average spread (max – min) is 0.067

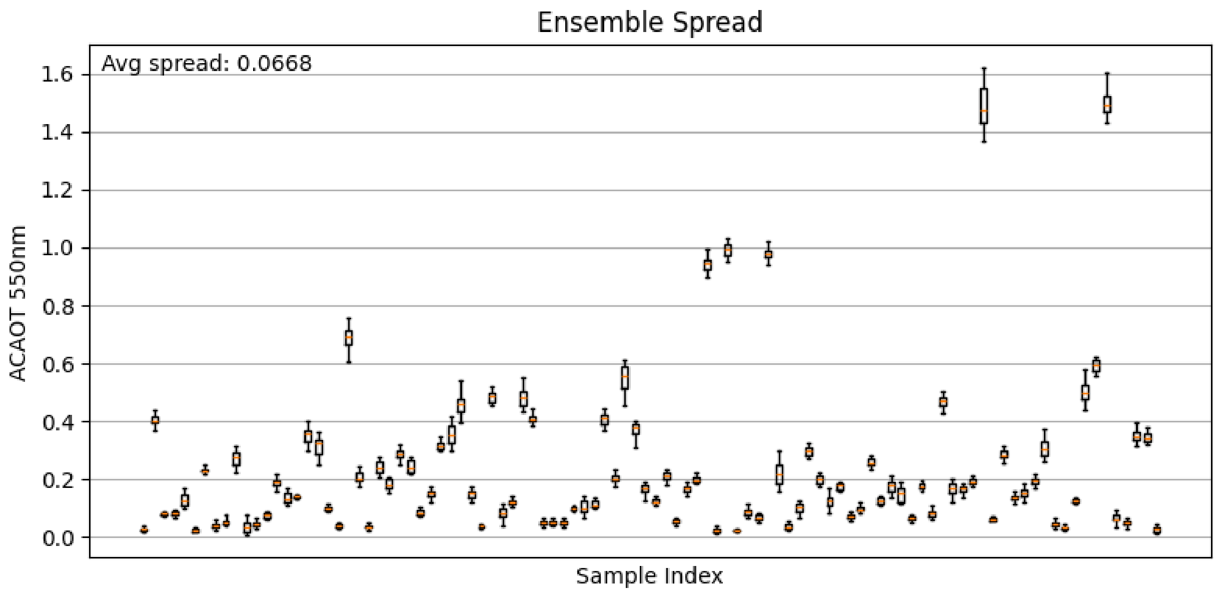

Fig 6. ACAOT (550nm) spread among 100 pixels (1%) on the synthetic validation set.

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.

One motivation for the smaller batch size (compared to other works) is to decrease the memory used in the training process. However, a large batch size benefits the convergence rate (Soham De, et al, 2017). We did several tests over batch size (from 512 to 20000) and didn’t find significant differences over the NN’s performance for our application. We added a related statement in the revised manuscript (line 189).

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.

We added more metrics in the plots and now it shows relative mean squared error (RMSE), mean absolute error (MAE), correlation coefficient (corr), number of pixels (npix) and coefficient of determination (R^2).

- ☐ 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.

We added MAE in the plots, but we believe relative MAE is not a good metric (especially for properties that can become close to zero) so it is not included. A table showing RMSE, MAE and bias of the synthetic experiments is included in SI (Table 2 of SI).

- ☐ 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.

It is possible that the NN performance may be improved for these extreme scenarios by adding more of such samples in the training set. We added a discussion on this aspect in the revised manuscript. (line 207)

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.

The ACAOT here means the ACAOT at 550nm, we clarified this (and also other ACAOT) in the revised manuscript.

- ☐ 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?

The liquid cloud fraction is a direct output of the 1st NN, which is at the original PARASOL resolution (6 km x 6 km). We also add this into the revised paper (line 215).

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

We added the wavelength to all the ACAOT in the revised paper.

☐ 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?

The wind speed variation have been taken into account in the training set. The geometry used to generate the training set is randomly taken from PARASOL real geometry (as is described in line 118 of the revised paper), which also include sun-glint areas.

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.

We have added additional metrics to the figures and a comparison with the PARASOL AERO-AC data product is included in the paper (see also our response above)

☐ 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.

We have added a case study in mid-Africa on 04 Aug 2008, showing above-cloud and clear-sky retrievals (see below). They are included in the paper.

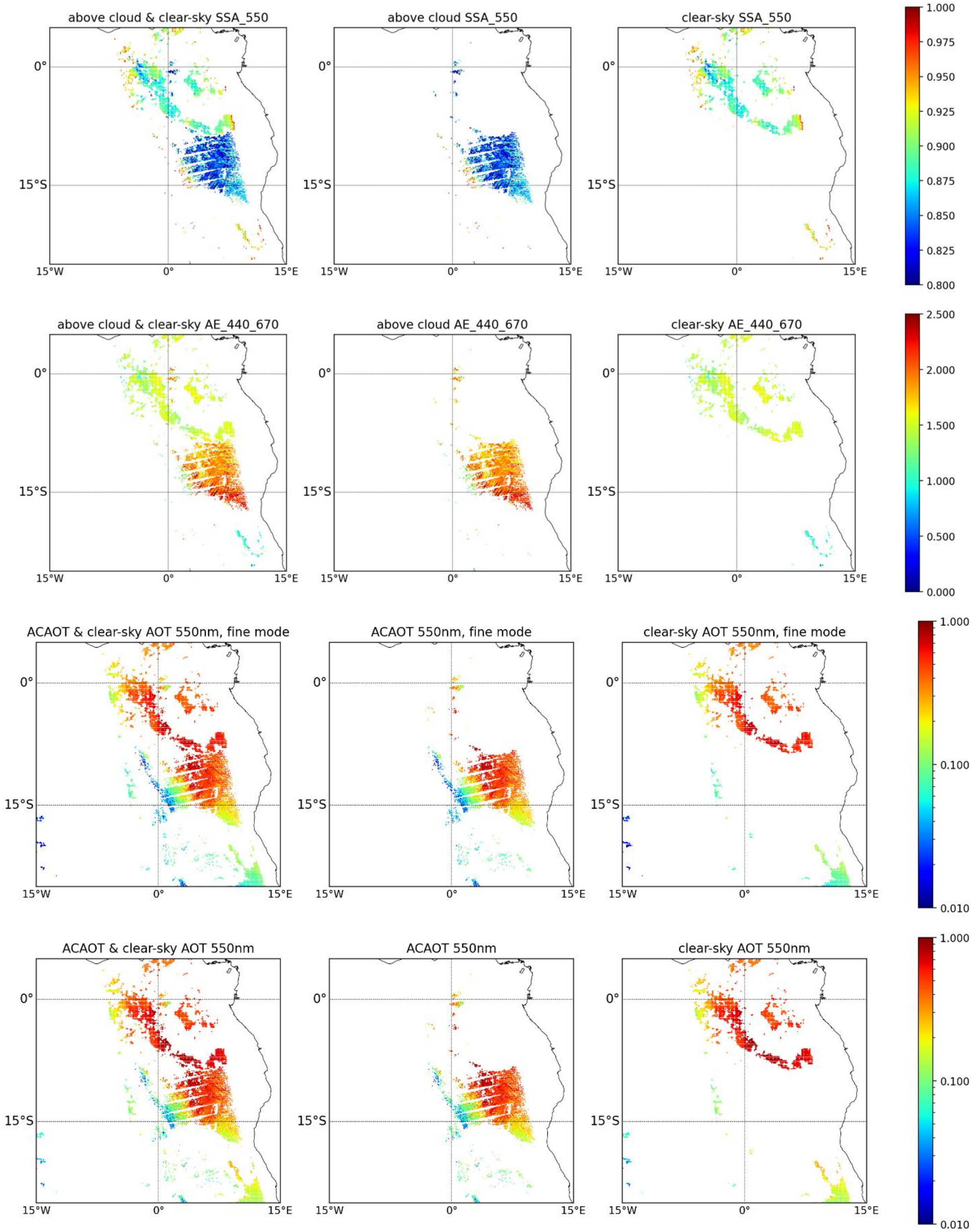

Fig 7. NN above cloud aerosol retrievals compared to RemoTAP clear sky aerosol retrievals in mid-Africa, 04 Aug 2008.

☐ 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?

We changed the colorbar of ACAOT plot to log-scale, and a histogram (as below) is included in the SI. The ACAOT is at 550nm.

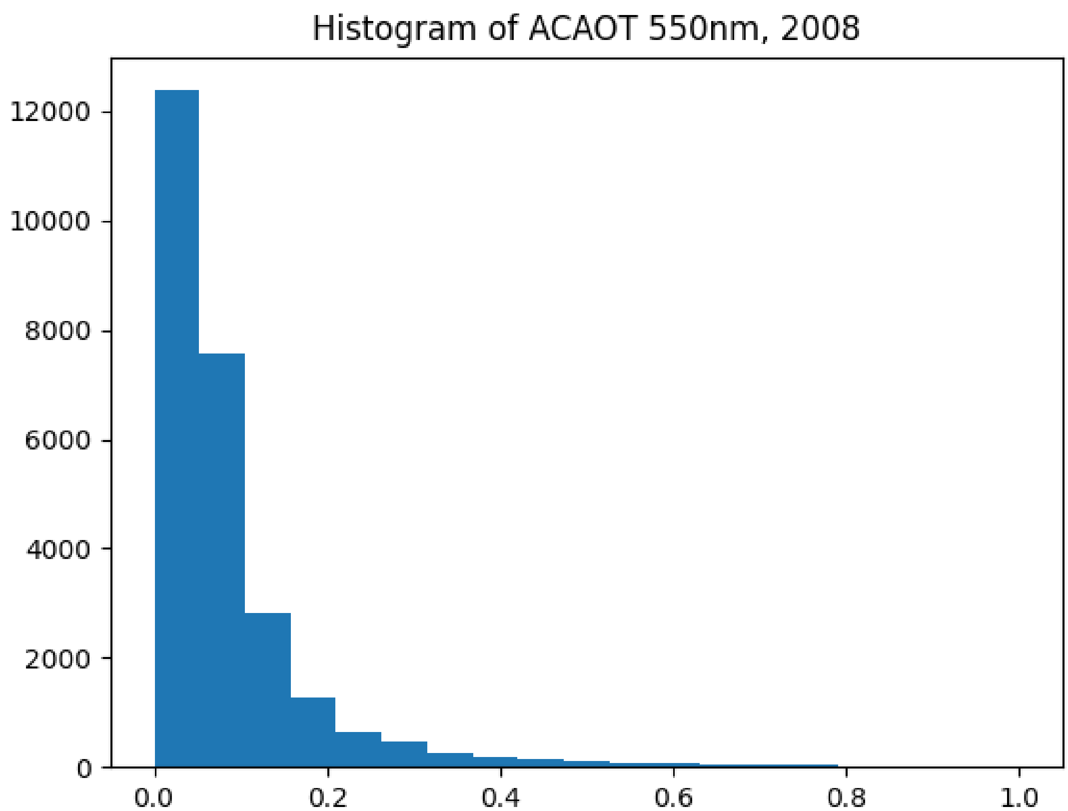

Fig 8. Histogram of ACAOT (550nm) for the whole year 2008 PARASOL-NN ACA retrievals.

☐ 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

We revised the paper based on this comment.

- ☐ 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.

We included another section (section 5.2) showing the comparison between PARASOL AERO-AC data product and the NN retrievals.

- ☐ 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

We only select where ACAOT_550 nm > 0.2 for plotting (and evaluating) AE of our PARASOL-NN The difference may be partly caused by low ACAOT cases in AERO-AC, but also in the direct comparison (including only larger ACAOT) we see much larger AE in AERO-AC. We added a discussion (section 5.2) of the revised paper.

- ☐ - 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?

Yes, the clear-sky retrievals retrieve a small contribution from the dust mode here, but overall the coarse aerosols are dominated by Sea Salt here. Both RemoTAP and GRASP give a low (clear-sky) AE in this region (Figure 10 in Hasekamp et al, 2024). The AE from that paper and this ACA paper are all filtered by AOT 550nm > 0.2.

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

  It is likely that the algorithm retrieves a contribution from the dust mode in the presence of coarse sea salt. But also dust may be present from e.g. Patagonia or Australia.