

## Comment on amt-2022-91

Anonymous Referee #3

---

Referee comment on "Performance evaluation for retrieving aerosol optical depth from the Directional Polarimetric Camera (DPC) based on the GRASP algorithm" by Shikuan Jin et al., Atmos. Meas. Tech. Discuss., <https://doi.org/10.5194/amt-2022-91-RC3>, 2022

---

The paper outlines an application of the GRASP algorithm to retrieval of aerosol optical depth from observations of the DPC multiangle polarimeter. Results are validated against the AERONET sun-photometer network, from which quality control metrics are devised, and a qualitative comparison is made to three MODIS aerosol products. The performance appears to be consistent with other remotely sensed aerosol products, which is impressive for a relatively new satellite and research team.

While I found the presentation generally good, I was disappointed by the meagre details provided by this manuscript. It would be impossible to replicate the method from this paper alone and the quantitative evaluation covers only AERONET. However, the authors have done a better job than many published works so I expect to see this work in print after some misunderstandings are corrected.

In the following, S means section and L means line number.

- At a glance, there is a substantial overlap between this paper and Li et al. 2022 as both apply GRASP to DPC. They evaluate different measurands and only share two authors, but my experience is that the AOD products discussed here are a by-product of the aerosol aerosol type products discussed there. The manuscript before me certainly provides additional information and I am not questioning the logic in publishing the projects separately. However, there must be more clarification of the relationship between the teams, either acknowledging how their work has complemented each other (e.g. I would hope that their determination of aerosol type provided the inputs to this method) or explaining why it was necessary to make separate implementations of the same code (to assist future GRASP users in determining which to use)?
- Throughout the paper the authors report Expect Error (EE%), being the number of retrieved values falling within some range of the validation value, and comment positively when this increases. Putting aside the fact that the authors never define the

term, nor state the envelope they use, this misunderstands the meaning of an error envelope. The MODIS error envelope is an estimate of a normally distributed error derived from comparison to validation data. As such, only 68% of data should fall within the error envelope (see "one sigma confidence interval"). Achieving a higher EE% does not mean the data is "better", merely that the EE overestimated the uncertainty in the circumstances considered. Ideally, the authors would estimate their own error envelope, which would presumably be narrower than that of MODIS. At a minimum, though, the authors must revise the language to express that the ideal EE% is 68%. (Also, Expected Error would be more grammatically correct.)

- While being clear that I don't expect the authors to change anything in the paper as they follow common and widespread practice, I will point out that the evaluation provided does not actually assess the accuracy of their retrievals. It assesses the accuracy of 30min/25km averages of their retrievals. Thus, the variability shown is a lower bound for the method's performance. This is clearly demonstrated in Fig.3(d), where accuracy improves as more observations are aggregated.
- The MODIS Dark Target, Deep Blue and MAIAC products are widely used, so I understand why the authors compare to them. But why do they not compare to MISR (or another polarimeter), which would provide a like-for-like comparison to another multiangle retrieval and demonstrate the relative merits of the GRASP method?
- I am more disappointed that, given the number of satellites it has been applied to, there is no comparison to another implementation of GRASP. That would provide valuable insight into the relative performance of the DPC sensor independent of retrieval method and assumptions.
- S2.3) Please be more precise as to the data used. Do you use every AERONET site in the record or do you exclude some? Do you report every collocation or do you exclude some? Do you use Level 1.5 or Level 2.0 as using both would be extremely foolish?
- L215-7) I do not know what you mean by 'absolute value of average relative deviations'. An equation would be clearer.
- S3.3) You don't appear to do any cloud filtering *before* averaging. Why not, given how common such approaches are in other aerosol retrievals?
- L236) The text states that the retrieval unit is 3x3 but Fig.1 shows a 5x5 unit. I appreciate that the larger cube helps illustrate the inclusion of different surface types in a single retrieval but please clarify what, precisely, is being done.
- Fig.2) I strongly agree with Dr. Kinne that this figure should be shown on a logarithmic scale. doi:10.5194/acp-19-15023-2019 provides compelling evidence that linear averaging of AOD provides misleading conclusions.
- L284) I disagree that Fig.2 shows that you underestimate AOD in high loading circumstances. I read that plot as showing an underestimate of AOD in typical circumstances (as the red blob is above the black line; see also the grey curve of Fig.4b). Your best-fit line has a gradient less than one largely because it's going through the peak of the distribution around 0.1 and towards the handful of points around 1.2.
- L286) There are numerous sources of error in any AOD retrieval and I would be surprised if the aging of the detector was the primary one.
- L297) On L255, you said that the method of external mixtures was under testing, which I took to mean "an experimental mode that will eventually be available". The text here implies that that is the mode you used. Please clarify what was done.
- L308) The retrieval residuals should conform to some distribution, such that very small values are not unexpected. Why is the polarized component different, requiring the exclusion of small residuals? My gut instinct would be a systematic bias in the observations or something about the representation of Rayleigh scattering in the forward model.
- Fig.6) This is an entirely qualitative comparison. There's nothing necessarily wrong with that, but I feel a page of description is inappropriate for two hand-picked scenes that cannot represent general performance. Perhaps the figure could appear at the start of the section as an illustration of the approach? Also, I feel the authors have

failed to mention the most important feature of this diagram: their method exhibits minimal land-sea contrast compared to others. This is a long-standing advantage of GRASP retrievals and limitation of other methods that is not widely acknowledged.

- Fig.7) While I appreciate an example of the performance over time, I do not like the units chosen (mean error ratio). As the text describes, MER can decrease both because a method has large, but complementary, errors; because the error envelope increases; or because different time steps present more/less difficult retrieval conditions. Further, L387 is not strictly correct as a lower MER can be achieved by a lower Error Envelope, which would happen if one method retrieves larger values than the other. More generally, I'm curious why the authors use normalised mean square error throughout the paper (without specifying by what the error is normalised; different communities would expect the retrieved value or the EE) rather than the more common root-mean square error?

Some more minor comments:

- L42) D'Almeida 1991 is a strange reference here, considering it's a microbiology paper.
- L126) By "normalized radiation", do you mean "reflectance"?
- L147) Please be specific what is meant by "highest quality" as different fields mean different things by it.
- L156) These are common collocation criteria, and I am not asking you to change anything here, but you may find it beneficial to read the series of papers Nick Schutgens has published on the best strategy to collocate different aerosol datasets, such as doi:10.5194/acp-20-12431-2020, 10.5194/acp-2015-973 and 10.5194/acp-16-1065-2016.
- L187) I disagree. It is entirely feasible to create a look-up-table-based method that integrates different instruments as, if you have an module that perform a calculation, it is possible to build a look-up table from it. The advantages of GRASP lie in its detailed radiative transfer simulations and multipixel approach. If you replace 'traditional look-up table-based methods' with 'most popular retrieval methods' I no longer have a problem with the sentence.
- L230) How often does DOLP>1 happen? Unless it's extremely rare, this filtering feels like it would introduce a low bias into that value.
- L259-60) Does 'exponential distribution' mean that your vertical levels are spaced exponentially, such that they are roughly equally spaced in pressure?
- L262) "General principles" is far too generic a description. Either remove it because the text that follows elaborates or explain what you mean.
- Fig.3) Please reproduce this figure so the text is a similar size to that in the figure's caption.
- L296) I think it would be clearer to refer to 'timesteps' rather than 'retrieval units'.
- L311) Also on L425. I think you've gotten the direction of the inequality wrong for the residuals, as it currently implies you remove high scattering angle and midline polarized residual. I also think you mean 'or' rather than 'and' as very few points will satisfy all of those conditions simultaneously.
- L414) It will certainly occupy an important position in China. The impact on the rest of the world will depend on the availability of the data.
- L671) While appreciating that the authors likely have little control over it, I see no reason for the nine-dotted line to appear in this plot as it is a political, rather than physical, boundary.

I attach an annotated PDF with typographic corrections I hope will be of use. They are largely verb tenses and use of 'the', which I certainly couldn't do accurately in another language. Red lines indicate text to delete while yellow highlight is for word replacement or insertion. As these were done by hand on a tablet, they do not cover precisely the words affected and all capitalisation in my comments should be ignored. Also, many of the citations either lack or incorrectly state the page number of the paper. If the authors have the time, it would be a massive improvement to include the DOIs of papers as this simplifies finding a paper.

Please also note the supplement to this comment:

<https://amt.copernicus.org/preprints/amt-2022-91/amt-2022-91-RC3-supplement.pdf>