## Interactive comment on "AEROCOM/AEROSAT AAOT SSA study, part I: evaluation and intercomparison of satellite measurements" by Nick Schutgens et al.

**Anonymous Referee #3**

The paper provides a much needed quantification of aerosol absorption properties derived by satellite products, by inter-comparison and comparison with AERONET. It contains sound scientific methods and statistics and with proper interpretation, this can help the satellite community and model community both to direct their efforts to improve aerosol absorbing properties.

My main criticism is a poor description of the algorithms in the method section, except for the POLDER-SRON part, and the lack of interpretation of the results. The issue of this paper is the accuracy of AAOD and SSA retrievals, or the absorbing properties of the aerosols. These are difficult parameters to retrieve from space, which has been

noted by the satellite community, and highlighted in this paper with a solid mathematical foundation. The paper's claim to be the first to show this may be correct, but there have been a number of papers in the past to show the minimum amount of information content that is needed before AAOD and SSA can be expected to be retrieved with some degree of accuracy. For POLDER-SRON a track record is available about the information content of multi-wavelength, polarisation and multi-angle measurements, which are essential for aerosol property sensing (e.g. Hasekamp 2005,2007). No discussion of the information content is given, while it would be quite interesting to know which information is combined for FL-MOC, or to understand that for the OMAERUV most of the information is assumed, while an instrument like POLDER was designed specifically to provide multi-angle, multi-wavelength and polarisation measurements.

A description of the information content that is available and used to extract the absorbing properties of aerosols in each algorithm is essential for the interpretation of the results. For the MOC (What is this? No description given) and OMI algorithms no information is given at all, only a reference to other papers. The reader should not have to resort to other papers to extract relevant information for this study. For the GRASP algorithm it should be made clear in what way it differs from the POLDER-SRON. At the very least, it should be clear what the information content is on which the derivation of the AAOD and SSA are based. In that way it can be evaluated whether only the algorithm difference (multi-pixels approach for GRASP or pixel-by-pixel for SRON) makes a difference, and that also different information content is the cause of the found differences.

This information should be included in the methods section, and a discussion should be added about the different results in terms of the expected accuracies that can be expected from the algorithms.

Connected to this issue is the question that the author pose themselves in section 2.1.6: How independent are these satellite products? This is a very important question, which is only very marginally answered, considering that two sets of data are based on

the same instruments (two based on OMI, two on POLDER). It correctly notices that the low correlation points to errors, not independence. It is essential for the understanding of the results that the different treatment of the (same) data are clear to the reader, so that the different results for different algorithms can be interpreted. Currently, the authors only present the errors or biases, but no explanation in terms of the algorithms' treatment of the different derivations of the AAOD and SSA.

Another issue I feel needs some discussion at least, if not answering, is the dependence on AERONET observations which is present in the OMAERUV and GRASP retrievals. OMAERUV incorporates a set of 24 aerosol models with microphysical properties that were based on AERONET observations. The subsequent selection of an aerosol type is based on 'geographical considerations..', which mostly means based on location. This means that over each AERONET station the OMAERUV aerosol microphysical properties are fixed and based on the AERONET observations themselves. For GRASP a similar bias towards the AERONET observations can be expected, since AERONET and GRASP have very similar inversion techniques and aerosol microphysical property assumptions. Size distributions and refractive indices are tuned to AERONET observations over the years. The GRASP inversion is 'versatile', but also so complex and based on many input parameters, weighing everything to get the best possible estimate of aerosol macro- and microphysical and surface properties. While this may be the best approach to get the best results, the independence of observations becomes questionable at best. A discussion of the independence of AERONET observations should be included here.

The text seems generally well written, but in places the authors are sloppy in the phrasing, grammatical structuring and the text. The paper is very technical, with increasing depths of mathematical treatment of the physics. The authors should take more care in guiding the reader in the statistical treatment. The paper treats the accuracy of AOD, AAOD and SSA derived from satellites. These are all connected parameters, but should not be interchanged, which seems sometimes the case.

E.g. in paragraph l56-l61 the authors state 'a number of satellite AAOD products have appeared,..'. Followed by: 'E.g. Lacagnina etal (2015) used POLDER data to evaluate SSA from AEROSOL models over oceans ; Peers etal (2016) estimated ocean above-cloud SSA in AEROCOM models.. ' Was POLDER AAOD used to evaluate SSA? Or would it be better to say 'a number of AAOD and SSA products have appeared'? Or was even more used in this paragraph? A more consistent grammatical structuring helps here so the reader can understand which parameters is meant.

The next paragraph is similar: in paragraph l62-l69 the authors describe the challenge in retrieving AAOD from satellite, using the challenge in retrieving AAOD from AERONET. In the remainder only SSA is mentioned, and the paragraph concludes that thus satellite SSA AND AAOD products can be expected to have large errors as well. This suggestion of a reasoning, with no logic following in the text, makes the paper hard to follow. Apparently, the reader has to conclude for him or herself that errors in SSA and AAOD are linked? However, the authors themselves conclude in the paper that SSA has much less skill than AAOD.

Another example:

l. 426. 'Over ocean, SSA products tend to correlate better than over land. The two POLDER products correlate better than any other satellite pair (r =âĹij 0.8 over ocean for AOD > 0.75).' The next paragraph starts like this: l 432. 'ost surprisingly, POLDER-GRASP-M and POLDER-SRON show a fairly systematic difference in SSA (-0.04), independent of AOD (there are regional variations).'

How are we to interpret these seemingly contradictory statements? Are we not talking about SSA? Are " two POLDER products" not the same as "POLDER-GRASP-M and POLDER-SRON" SSA? Or are the statements not contradicting? Probably the latter, but the reader has to check his/her own sanity a couple of times first, before this become apparent. In a technical paper like this consistent phrasing and grammatical structuring is even more essential than normally, and the lack of that in the current

paper makes it hard to read.

Other minor issues are listed below:

2 l35 heating can also destabilise the boundary layer (Johnson et al, 2004), semi-direct effect are now called fast adjustments and can be both negative and positive in forcing.

l158 The terms bidirectional reflectance distribution function (BRDF) and bidirectional polarisation distribution function (BPDF) are not explained.

Figure 4. Half of the difference plots are the same (but vv) and can be removed.

l305 "The scatter plots show good correlation with AERONET." This is a meaningless term. The idea is to quantify the goodness, or accuracy. Please, rephrase to The scatter plots show the correlation of the satellite AOD with AERONET AOD.

l326 product -> products l346 Section -> section l349 0.006

l. 365 If cloud contamination if such a big problem, why is it not (additionally) removed?

Figure 12. For POLDER-GRASP-M an additional minimum AOD threshold is used before calculating AAOD and aggregating SSA (l 166.). The threshold is not mentioned in the paper. However, it is not 0, as suggested in the caption of Fig 12. This should be clear in the Figure and/or the text.

l 432. 'Most surprisingly, POLDER-GRASP-M and POLDER-SRON show a fairly systematic difference in SSA (-0.04), independent of AOD (there are regional variations). A major exception would be cases over the deep ocean at low AOD (< 0.1) where this bias dissappears.'

Is this not a result from the fact that no absorbing aerosols are left over the 'deep' oceans? I expect deep oceans refer to those remote parts far from the land (thus aerosol sources), where only clouds and marine aerosols are left ? One would not expect any signal left for those areas. In that case it would make sense refer to 'remote oceans' or something.

dissappears -> disappears

l 440. 'than is present in' -> compared to

l 442. 'It will not be easy to increase Inversion L2.0 observations' -> "It will not be easy to increase THE NUMBER OF Inversion L2.0 observations"?

Johnson, B.T., Shine, K.P. and Forster, P.M. (2004), The semi-direct aerosol effect: Impact of absorbing aerosols on marine stratocumulus. Q.J.R. Meteorol. Soc., 130: 1407-1422. https://doi.org/10.1256/qj.03.61

Hasekamp, O. P., and J. Landgraf (2005b), Retrieval of aerosol propertiesover the ocean from multispectral single-viewing-angle measurements ofintensity and polarization: Retrieval approach, information content, andsensitivity study, J. Geophys. Res., 110, D20207, doi:10.1029/2005JD006212.

Hasekamp, O. P., and J. Lan dgraf (2007), Retrieval of aerosol propertiesover land surfaces: Capabilities of multiple-viewing-angle intensity andpolarization measurements, Appl. Opt., 46, 3332–3344.