Reviewer #3:

Dear Greg,

We would like to thank you for the constructive and positive comments on our paper. Please find our responses below.

This paper applies several versions of a relatively new retrieval algorithm (GRASP/HP, GRASP/Optimized, GRASP/Models) to an existing satellite measurement archive (POLDER/PARASOL). The authors then compare the results from these algorithms to several legacy retrieval algorithms, including MODIS Dark Target (MODIS/DT), Deep Blue (MODIS/DB), Multi-Angle Implementation of Atmospheric Correction (MAIAC), the operational PARASOL product (PARASOL/Operational), and AERONET. The authors provide a large number of maps and statistics that not only inform the reader about the performance of the GRASP products, but they also inform the reader about the performance of the legacy aerosol retrievals (e.g., Tables 9 & 10). The paper is clear and well written and I find it suitable for publication.

Response:

Thank you for the positive comments on our manuscript. We provide point-by-point responses as follows.

The paper is also quite long (48 pages of text, 23 tables, and 28 figures) and probably won't be carefully read in its entirety by anyone except the reviewers. One could paraphrase the paper as "here are some new data products, and here is how they compare to similar data products as well as the gold standard (AERONET)." Nobody will learn the machinery behind the retrievals from this paper, but there are other papers for that. One reviewer pointed out that the statistical parameters chosen for this paper are not ideal, but the authors use statistical parameters that are familiar to many readers (correlation coefficient, bias, RMSE, etc.) and common in many satellite/AERONET comparison papers. Unfortunately, the aerosol remote sensing community has not yet adopted a "skill score" for comparisons, as is sometimes used in the modeling community (Taylor, 2001).

It is important to have all of this material in one place, in my opinion, so that readers can quickly assess the relative performance of the different algorithms for the various

parameters. However, hyperlinks to tables, figures, citations, and section headings would greatly improve the readability of the paper. A bookmarked Table of Contents in the sidebar would also be helpful. I had to keep two copies of the paper open on my screen – one for the text, and another for reading tables and figures. Otherwise, I would have spent as much time scrolling as reading for this paper! Hyperlinks would allow the reader to go directly to a table, and then return to the text with the "previous view" buttons. Hopefully this is something that can be accommodated in the typesetting process.

Taylor, K.: Summarizing multiple aspects of model performance in a single diagram, J. Geophys. Res., 106, 7183–7192, 2001.

Response:

These are very constructive suggestions. We have revised the manuscript and included a Table of Content and hyper link of all tables and figure.

I noticed that the data volume for GRASP/Models is much different than the data volume for GRASP/HP and GRASP/Optimized. For instance, GRASP/Optimized shows 41,268 AOD comparisons with AERONET over land in Table 3, but GRASP/Models only shows 27,551 comparisons. However, GRASP/Model comparisons are greater than GRASP/Optimized over ocean (2064 vs 1495). These large discrepancies appear elsewhere in the paper as well. I found this quite odd, since all three retrievals use the same instrument (PARASOL). I imagine that the cloud screening procedure is identical for all three algorithms, so I suspect that GRASP/Model fails to provide a retrieval much more frequently over land than than the other two GRASP algorithms (and that GRASP/Optimized, GRASP/HP fail more frequently than GRASP/Models over ocean). This should be discussed, since GRASP/Models is lauded for its ability to retrieve AOD (550) (e.g., line 1180). The success rate of a retrieval is important to readers, too.

Response:

Indeed, the different number of points and somewhat different approaches of quality filtering is one of the main shortcomings in our study. Reviewer #1 raised similar question. In fact, the post-processing flow (L1-L2-L3) was based on several attempts dictated by practical needs. These attempts provided us valuable inside but the could be fully evaluated only after full-scale validation. For example, the level 3 GRASP/HP and Optimized archives were generated and released much earlier than

GRASP/Models. Also, we have done preprocessing of GRASP/Models over land at first and learned that screening was very conservative. Based on that we used less conservative screening for Models reprocessing over ocean. Once the products were released, they were used by many users, therefore, regenerating Level 3 products was not reasonable for this study. We are considering the harmonization of the all archives in future once time and resources allow that.

At the same time, we have looked at possible effect of applying tighter screening to HP and Optimized data. Our analysis showed that although stricter screening somewhat improves the correlations, it doesn't change conceptually the results of validations. For example, it does not improve the BIAS, which is considered as a main issue for these data sets. Some explanations of this aspect were added in the Sect 2.4 and Sect 3.1 as follow.

It is also curious that GRASP/Models did so well for AOD at nearly all wavelengths over both land and ocean (Table 3), but the AEs for GRASP/Models is significantly worse than the other GRASPs. Since AE is derived from AOD, I would have thought the retrieval that produced the best AOD at multiple wavelengths would also produce the best AE. A comment about this would be helpful.

Response:

This is a good point that should be mentioned in the text. Yes, according to the analysis, we found out that the good agreement of spectral AOD is not equivalent to the good agreement for AE, even though AE is derived spectral AOD at two wavelengths. Apparently, obtaining good agreement for spectral AOD seems easier than for AE. The potential reason is that the spectral contrast is crucial to calculate AE, and small error in AOD at each wavelength can result in large AE uncertainty. The level of uncertainty (e.g. ±0.1, RMSE=0.1~0.15) of satellite-derived spectral AOD may makes it challenging accurate estimation of AE. At the same time, the same uncertainty is sufficient for good agreement of AOD at different wavelengths. Thus, the AOD in each channel may correlate well in time, while for each singe retrieval spectral deviations can be significant. Also, relatively small spectral deviations may perturb AE strongly while not to be as notable for AOD at each wavelength.

Is there a reason for comparing AODf and AODc to the SDA extinction-based retrievals instead of using the sky scan retrievals? The sky scans are probably more accurate. Many readers (most?) won't know the methodology behind SDA and may incorrectly assume that it is derived from the sky scans. The SDA papers use the AERONET almucantar scans as a performance benchmark, so why not use the same benchmark? You're already using the AERONET sky-scans for SSA and AAOD.

Response:

We have done AODF and AODC comparison with AERONET ALM retrievals (see in Figures R2 and R3). In general, they show quite similar performance with Figures 4 and 5 that evaluation against AERONET SDA AODF and AODC. Please note, in order to find more matched pairs, the AERONET ALM retrievals are collocated within ±180mins for satellite overpass, which is much bigger time window that that for SDA products ±30mins. To ensure the retrieval quality, AERONET ALM L2 retrieval products are available for AOD440>0.4, which roughly filter 80% low AOD cases (in future we could use AERONET L1.5 inversion products). By using SDA extinction-based products, we get almost the same amount of points for AOD, AODF, and AODC, which help to understand the overall performance for low, medium and high AOD cases.

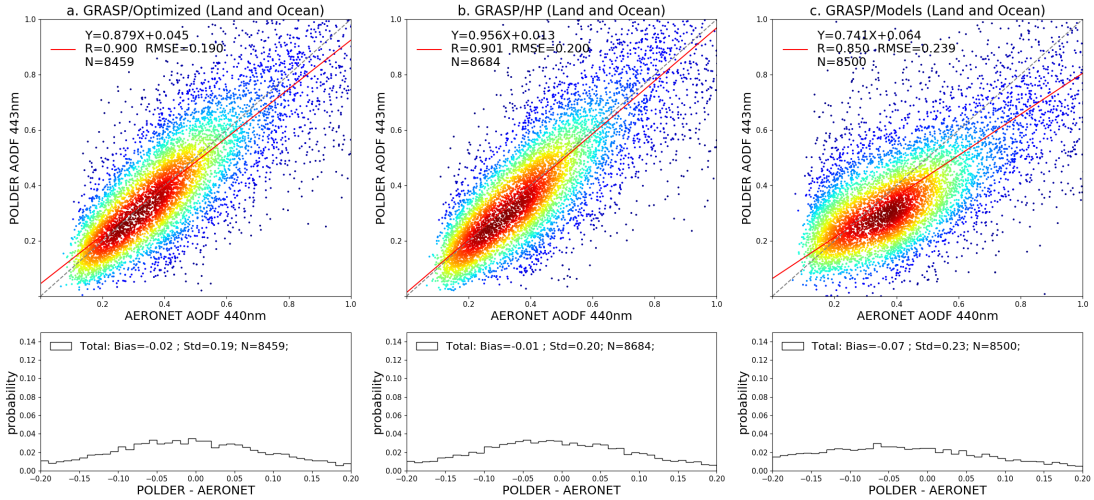

Figure R2. Evaluation of all archive PARASOL/GRASP AODF (2005-2013) at 440 nm with AERONET INV AODF, (a) GRASP/Optimized; (b) GRASP/HP; (c) GRASP/Models.

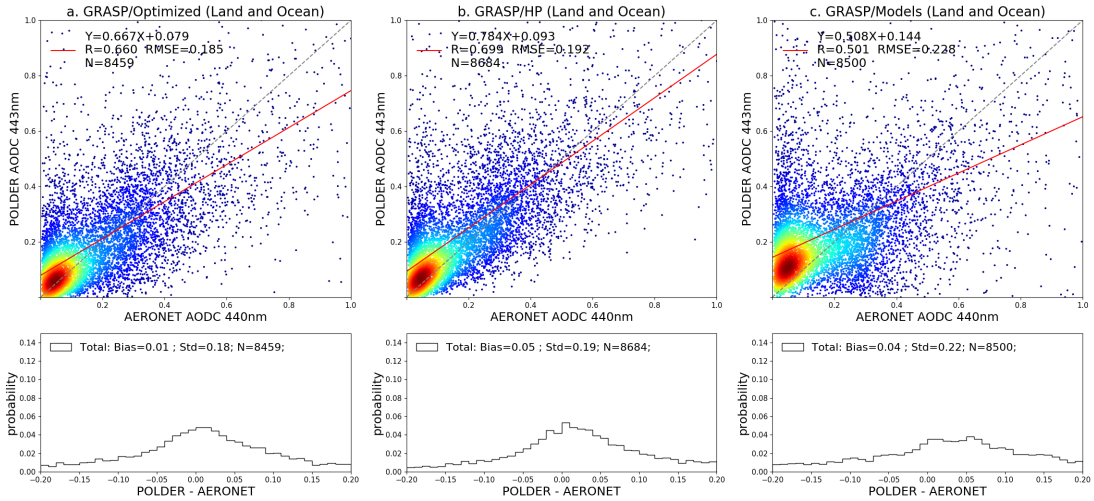

Figure R3. The same with Figure R2, but for AODC at 440 nm

Line 588:

The authors say that they are using AERONET L2 inversions, but which version of AERONET (i.e., Version 2 or Version 3)?

Response:

We are using Version 3 AERONET L2 inversion. We now clarify this in the text.

Lines 590-595:

This paragraph will probably confuse some people. Line 590 says that AERONET L2 inversion products require AOD(440) > 0:4, but the PARASOL/GRASP filtering includes much lower values, especially over ocean (the authors require PARASOL/GRASP AOD(443) > 0:3 over land and AOD(443) > 0:02 over ocean). However, since AERONET L2 requires AOD(440) > 0:4, many of the low PARASOL/GRASP AODs won't actually appear in the comparisons anyways. . . . Unless the authors using Level 1.5 AERONET inversions at low AOD, like some other authors? If so, what are the Level 1.5 constraints?

Response:

We are using Version 3 AERONET L2 inversion products, which includes only AERONET AOD>0.4. Here we use additional filter for satellite AOD (Land: AOD 443 nm>0.3; Ocean: AOD 443 nm>0.02).

MINOR ISSUES

Line 35:

The links do not take me directly to the data products. www.icare.univ-lille.fr takes me to the main page, and the 2nd link on that line tries to take me to www.grasp-35, but that is a dead end.

Response:

Yes, the provided link is to ICARE main page, the PARASOL products are published at this path: https://www.icare.univ-lille.fr/data-access/data-archive-access/?dir=PARASOL/, ICARE account, that is free registration, is required to login. For the second link, it should work as https://www.grasp-open.com/products/.

Line 1125

Do you mean GRASP/Models instead of PARASOL/Models?

Response:

Yes, we use 'PARASOL/Models' to represent PARASOL products generated by GRASP/Models approach.

It would be interesting to repeat the AOD comparisons using AERONET's "coincident" AOD (that is, using only the AODs that are used during the sky scans). This would be interesting because the cloud screening for the sky-scan products is more comprehensive than for the direct AOD measurements, and it is possible that satellite (and model) AOD performance comparisons wrt AERONET will differ for these two datasets. If the coincident AODs comparisons are different than the "all AODs" comparisons, this could assist our thinking wrt the other sky-scan products. You are probably already set up to do this. I include this as a minor issue, though, because the paper is already too long and this should really be a topic for another day.

Response:

This is a good suggestion, which brings us additional thoughts! We agree that AERONET direct sun cloud screening is not as comprehensive as for sky-scan products. However, the evaluation with collocated satellite retrievals also introduces the satellite cloud screen to ensure the quality. On the other hand, when utilizing AERONET sky-scan products, in order to find more matched pairs we normally adopt a wider time window, e.g. ±180 mins, which may increase the issue of aerosol temporal variability.