Reviewer #1:

The paper describes a thorough evaluation of aerosol products retrieved by the GRASP algorithm (in different configurations) from POLDER-3/PARASOL. First, a comparison to AERONET for the full data set is presented. Second, a comparison with MODIS aerosol products is performed. It is concluded that the GRASP/Models AOD product is at least as good as (and probably better than) the MODIS AOD products and that the GRASP/HP product is superior for retrieving SSA and AE.

Overall, the paper is well written and the conclusions are sound. The part on the comparison with MODIS is quite detailed and sometimes a bit hard to follow (because of the comparison of 3 GRASP products with 3 MODIS products). I think this part can be shortened by removing the part of fine- and coarse mode AOD as I believe the AOD+AE comparison already tells the story.

I recommend publication of this paper after addressing my comments I added to the pdf file of the manuscript, most of which are minor.

Response:

Dear Otto,

Thank you for the constructive and positive comments on our paper. Here are the point-by-point responses:

We agree that the Ångström exponent (AE) can qualitatively indicate the domination of fine or coarse model aerosol. At the same time, we would like to point out that quantitative analysis of fine/coarse mode contribution to total AOD is very challenging and unclear without fine mode AOD (AODF) and coarse mode AOD (AODC). Moreover, the AODF is a parameter of particular interest for various applied research targeting characterization of Air Quality and anthropogenic effects. For example, the Air Quality related studies of Wei et al. (2020) are focused on only AODF from POLDER and MODIS. In addition, the multi-angular polarimetry is known for high sensitivity to fine mode aerosol and the POLDER/Operational products provide AODF over land and ocean as the main "operational POLDER" products. Therefore, we consider beneficial to keep AODF and AODC analysis in our paper. We validated the POLDER/GRASP AODF with AERONET AODF, and inter-compared with POLDER/Operational and MODIS over ocean. We believe that our analysis provide useful insights for the users of satellite AODF and AODC products.

Wei, Y., Z. Li, Y. Zhang, C. Chen, O. Dubovik, Y. Zhang, H. Xu, K. Li, J. Chen, H. Wang, B. Ge, C. Fan, "Validation of POLDER GRASP aerosol optical retrieval over China using SONET observations", J. Quant. Spectrosc. Radiait. Transfer, 246, 106931, 2020. doi:10.1016/j.jqsrt.2020.106931, 2020.

Main points:

Two comments I'd like to highlight here: - It seems that the GRASP/Models product has significantly less valid retrievals than the GRASP/HP product (~31000 vs ~44000). What is the reason? Is the filter for GRASP/Models stricter? This is not clear from the text (in fact the opposite is suggested). May this be the reason for the better performance? Some discussion is needed here.

Response:

Indeed, the different number of points and somewhat different approaches of quality filtering is one of the main shortcomings in our study. Reviewer #3 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 they 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.

"*For GRASP/Models product we did not use any filter, because a stricter quality assurance filter has been applied in GRASP/Models products generation from L1 to L2 and L3 than for other GRASP datasets. In principle, the post-processing of all PARASOL/GRASP products was done in similar ways. At the same time, the L3 products were prepared and released not at the same time. For example, the L3 GRASP/HP and Optimized archives were generated and released much earlier than GRASP/Models. Therefore, the post-processing and quality screening approaches used for different data archives are not exactly the same. Unfortunately, most of the differences were identified after the release of the products, its extensive use and the full-scale validation. In these regards, the harmonization of the all archives is likely to be done in future, but it will lead in release of new products.*"

"*As described in section 2.4, the different post-processing scheme resulted in the difference for matched points between GRASP/Optimized, GRASP/HP and GRASP/Models. It can be noticed that applying a much stricter filter may improve the overall correlation against AERONET AOD for GRASP/HP and GRASP/Optimized products, but leads to significant loss of points and, most importantly, do not improve the BIAS, which is considered as a main issue for them.*"

The evaluation puts large focus on the correlation coefficient when comparing the performance of different products. This is not always a good metric because it is heavily influenced by the range, i.e. a limited number of points at the end of the range can have large effect on the correlation. I recommend to put more emphasis on other metric such as RMSE and MAE (Mean Absolute Error).

Response:

The idea of our analysis was to provide comprehensive characterization of the observed relationships, rather than to focus on one selected parameter. Therefore, we provided many different parameters describing the relationship and did not intend to focus on correlation coefficient. Thanks for the suggestion about evaluation metrics! We have revised the text to make more comprehensive discussion of other correlation parameters including RMSE and BIAS.

At the same time, we would like to note that RMSE and BIAS also have limitations in characterizing quality of agreement. For example, RMSE is always much lower for the dataset that include mainly small values of AOD. Therefore, the retrievals failing to report the retrieval at larger AODs have smaller RMSE than those that provide AODs with larger values.

Minor points and grammar:

Line 103: Please also cite Mishchenko and Travis, JGR, 1997 doi:10.1029/96JD02425 and Hasekamp and Landgraf, Appl. Opt., 2007, https://doi.org/10.1364/AO.46.003332)

Response:

Corrected.

Line 139: 910 nm

Response:

Corrected.

Line 159: Mention that also other MAP algorithms have been developed / are being devloped: SRON (Hasekamp et al., J. Geophys. Res., 116, D14204, https://doi.org/10.1029/2010JD015469, 2011; Fu and Hasekamp https://doi.org/10.5194/amt-11-6627-2018, 2018) JPL (Xu et al., https://doi.org/10.1002/2017jd026776, 2017; https://doi.org/10.3390/rs11070746, 2019), LaRC (Stamnes et al., https://doi.org/10.1364/AO.57.002394, 2018.) All these algorithms follow a similar principle of online RT calculations and no restriction to aerosol models.

Response:

We fully agree that the text of article didn't mention other advanced algorithms and therefore could be misleading. We added corresponding discussion in Sect. 2.1 as below.

"*A unique aspect of GRASP is that it can perform radiative transfer (RT) computations fully accounting for multiple interactions of the scattered solar light in the atmosphere online without the use of traditional LUTs. Several other algorithms of new generation have been or being developed for interpretation of MAP observation use the online RT calculations and implement retrieval as a search in* *continuous space of solution e.g. Hasekamp et al., 2011, Xu et al., 2017, 2019, Fu and Hasekamp, 2018, Gao et al., 2018, Stamnes et al., 2018, Di Noia et al., 2019. Nonetheless, at present GRASP was the only algorithm that has been used to generate aerosol products for full archive of POLDER observations (Dubovik et al., 2019).*"

Dubovik, O., Li, Z., Mishchenko, M. I., Tanré, D., Karol, Y., Bojkov, B., Cairns, B., Diner, D. J., Espinosa, W. R., Goloub, P., Gu, X., Hasekamp, O., Hong, J., Hou, W., Knobelspiesse, K. D., Landgraf, J., Li, L., Litvinov, P., Liu, Y., Lopatin, A., Marbach, T., Maring, H., Martins, V., Meijer, Y., Milinevsky, G., Mukai, S., Parol, F., Qiao, Y., Remer, L., Rietjens, J., Sano, I., Stammes, P., Stamnes, S., Sun, X., Tabary, P., Travis, L. D., Waquet, F., Xu, F., Yan, C. and Yin, D.: Polarimetric remote sensing of atmospheric aerosols: Instruments, methodologies, results, and perspectives, J. Quant. Spectrosc. Radiat. Transf., 224, 474–511, doi:10.1016/J.JQSRT.2018.11.024, 2019.

Fu, G. and Hasekamp, O.: Retrieval of aerosol microphysical and optical properties over land using a multimode approach, Atmos. Meas. Tech., 11(12), 6627–6650, doi:10.5194/amt-11-6627-2018, 2018.

Gao, M., Zhai, P.-W., Franz, B., Hu, Y., Knobelspiesse, K., Werdell, P. J., Ibrahim, A., Xu, F. and Cairns, B.: Retrieval of aerosol properties and water-leaving reflectance from multi-angular polarimetric measurements over coastal waters, Opt. Express, 26(7), 2973–2984, doi:doi.org/10.1364/OE.26.008968, 2018.

Hasekamp, O. P. and Landgraf, J.: Retrieval of aerosol properties over land surfaces: Capabilities of multiple-viewing-angle intensity and polarization measurements, Appl. Opt., 46(16), 3332–3343, doi:10.1364/AO.46.003332, 2007.

Hasekamp, O. P., Litvinov, P. and Butz, A.: Aerosol properties over the ocean from PARASOL multiangle photopolarimetric measurements, J. Geophys. Res., 116(D14), D14204, doi:10.1029/2010JD015469, 2011.

Di Noia, A., Hasekamp, O. P., Van Diedenhoven, B. and Zhang, Z.: Retrieval of liquid water cloud properties from POLDER-3 measurements using a neural network ensemble approach, Atmos. Meas. Tech., 12(3), 1697–1716, doi:10.5194/amt-12-1697-2019, 2019.

Stamnes, S., Hostetler, C., Ferrare, R., Burton, S., Liu, X., Hair, J., Hu, Y., Wasilewski, A., Martin, W., van Diedenhoven, B., Chowdhary, J., Cetinić, I., Berg, L. K., Stamnes, K. and Cairns, B.: Simultaneous polarimeter retrievals of microphysical aerosol and ocean color parameters from the "MAPP" algorithm with comparison to high-spectral-resolution lidar aerosol and ocean products, Appl. Opt., 57(10), 2394, doi:10.1364/ao.57.002394, 2018.

Xu, F., van Harten, G., Diner, D. J., Kalashnikova, O. V., Seidel, F. C., Bruegge, C. J. and Dubovik, O.: Coupled retrieval of aerosol properties and land surface reflection using the Airborne Multiangle SpectroPolarimetric Imager, J. Geophys. Res., 122(13), 7004–7026, doi:10.1002/2017JD026776, 2017.

Xu, F., Diner, D. J., Dubovik, O. and Schechner, Y.: A correlated multi-pixel inversion approach for aerosol remote sensing, Remote Sens., 11(7), doi:10.3390/rs11070746, 2019.

Line 188: SPEXone

Response:

Corrected.

Line 191: It is important to mention in the introduction that advanced MAP aerosol products from POLDER-3 (from GRASP and SRON algorithms) have already been used for emission estimates (Chen et al, https://doi.org/10.5194/acp-18-12551-2018, 2018; https://doi.org/10.5194/acp-19-14585-2019, 2019); data assimilation (Tsikerdekis et al., https://doi.org/10.5194/acp-2020-468), estimating the direct radtive effect of aerosols (Lacagnina et al., https://doi.org/10.1002/2015jd023501, 2015. https://doi.org/10.1002/2016jd025706, 2017) and radiative forcing due to aerosol-cloud interactions (Hasekamp et al., https://doi.org/10.1038/s41467-019-13372-2, 2019).

This shows the potential of such advanced MAP products.

Response:

We now provide discussion as follow.

"*Several studies have shown the potential of advanced MAP aerosol products, for example, PARASOL/GRASP results have been adopted to estimate global aerosol emissions (Chen et al., 2018; 2019), PARASOL/SRON products have been used for data assimilation (Tsikerdekis et al., 2020), estimation of aerosol direct radiative effect (Lacagnina et al., 2015; 2017), and the radiative forcing due to aerosol-cloud interactions (Hasekamp et al., 2019b).*"

Chen, C., Dubovik, O., Henze, D. K., Chin, M., Lapyonok, T., Schuster, G. L., Ducos, F., Fuertes, D., Litvinov, P., Li, L., Lopatin, A., Hu, Q. and Torres, B.:

Constraining global aerosol emissions using POLDER/PARASOL satellite remote sensing observations, Atmos. Chem. Phys., 19(23), 14585–14606, doi:10.5194/acp-19-14585-2019, 2019.

Chen, C., Dubovik, O., Henze, D. K., Chin, M., Lapyonok, T., Schuster, G. L., Ducos, F., Fuertes, D., Litvinov, P., Li, L., Lopatin, A., Hu, Q. and Torres, B.: Constraining global aerosol emissions using POLDER/PARASOL satellite remote sensing observations, Atmos. Chem. Phys., 19(23), 14585–14606, doi:10.5194/acp-19-14585-2019, 2019.

Hasekamp, O. P., Gryspeerdt, E. and Quaas, J.: Analysis of polarimetric satellite measurements suggests stronger cooling due to aerosol-cloud interactions, Nat. Commun., 10(1), 1–7, doi:10.1038/s41467-019-13372-2, 2019b.

Lacagnina, C., Hasekamp, O. P., Bian, H., Curci, G., Myhre, G., van Noije, T., Schulz, M., Skeie, R. B., Takemura, T. and Zhang, K.: Aerosol single-scattering albedo over the global oceans: Comparing PARASOL retrievals with AERONET, OMI, and AeroCom models estimates, J. Geophys. Res., 120(18), 9814–9836, doi:10.1002/2015JD023501, 2015.

Lacagnina, C., Hasekamp, O. P. and Torres, O.: Direct radiative effect of aerosols based on PARASOL and OMI satellite observations, J. Geophys. Res., 122(4), 2366–2388, doi:10.1002/2016JD025706, 2017.

Tsikerdekis, A., Schutgens, N. A. J. and Hasekamp, O. P.: Assimilating aerosol optical properties related to size and absorption from POLDER/PARASOL with an ensemble data assimilation system, Atmos. Chem. Phys. Discuss., in review, doi:10.5194/acp-2020-468, 2020.

Line 210: There are also other algorithms that do that (see my comment above).

Response:

Corrected.

Line 307: In Introduction together with other MAP aerosol applications (see above).

Response:

Corrected.

Line 416: I would say for SSA/AAOD it is more comparison than validation because also AERONET has large uncertainties for these properties.

Response:

We agree that SSA and AAOD provided by AERONET have substantial uncertainties in some situations (such as the low-AOD case). At the same time, we consider that AERONET, so far, provides overall the most comprehensive and reliable SSA and AAOD data among all available data sources. Since AERONET retrievals rely on direct Sun observation for getting AOD, they have serious advantages over satellite data for constraining aerosol absorption retrieval.

At the same time, we do not consider our POLDER paper as a right place for the discussions, we rewrote the sentence as follows to avoid the discussion: *"AAOD and SSA products are chosen as references for satellite products comparison and evaluation."*

Line 430: How to go from 6km to 0.1 degree?

Response:

The gdalwarp regridding technique (https://gdal.org/programs/gdalwarp.html) is used to generate 0.1 degree products from original ~6 km retrieval. The algebraic mean is then calculated from all pixels in a grid box.

Line 440: How to interpret these numbers (residual)?

Response:

The relative residual is the root mean square of relative error in fitting the measurements by the algorithm.

Line 441: Not clear. Actually the GRASP/Models results seem most heavily filtered as there are much less points in the validation plots.

Response:

Rather elaborated filtering scheme was used to generate GRASP/Models L3 products. Hence, no additional filtering was applied to L3 GRASP/Models 0.1degree products in the validation considerations.

Line 450: So, this is 9 0.1 degree pixels?

Response:

Not exactly. For POLDER/GRASP and MODIS DT, DB and MAIAC_0.1, it is 9 0.1 degree pixels. While for MAIAC_0.01, 9 0.01 degree pixels are used, and 9 18.5 km pixels are for POLDER/Operational products. We now clarify this in the text.

Line 461: Why? Indications for clouds?

Response:

The purpose of this criterion is to remove some evident outliers, which stand out within 3x3 or 9x9 windows. It is proven to be helpful for validation. It is quite possible that the high variability in 3x3 or 9x9 windows can be related to clouds, but further investigations for supporting this idea are needed.

Line 480: I would say the AERONET uncertainty should be added quadratically ....

Response:

We agree that if we consider the statistical rule of adding standard deviations the quadratic addition should be done. However, GCOS criteria seem to be a criterion defined based on practical considerations rather than rigorous statistical consideration. In this regards, we simply followed the common practice and have adopted the GCOS requirements, $\mathrm{GCOS} = \max(\pm 0.04, \pm 0.1\mathrm{AOD})$,following the latest Aerosol_cci study (Popp et al., 2016).

Line 489: Why just 2008?

Response:

The year 2008 was chosen as an example year to comprehensive evaluation of the consistencies and differences between POLDER/GRASP and MODIS aerosol products. The observations during 2008 year contain generally good observational statistics and all types of aerosol are clearly present. The year has also been used as a reference in many evaluation studies, e.g., in Aerosol_cci study.

Line 505: But there are much less points.

Response:

Yes, this is correct, GRASP/Models product contains less point than GRASP/Optimized and GRASP/HP due to different filter scheme used for generation of L3 0.1degree products. This aspect has already been discussed above and a discussion is added in the Sect. 2.4 and Sect 3.1 of the revised manuscript.

Line 543: Do fine and coarse mode have the same definition on GRASP and AERONET?

Response:

The separation of the fine and coarse modes in AERONET and GRASP are not exactly the same, but they provide very close results in case then both modes are separated by distinct minimum.

Line 582: So, is GRASP/Models really better? Or is it a 'lucky' compensation of errors?

Response:

In terms of spectral AOD products, GRASP/Models show better agreement with AERONET measurements in many senses. The biggest problem of GRASP/Optimized and GRASP/HP AOD products is a distinct BIAS. At the same time, the validation of AODF and AODC indicates that the BIAS for GRASP/Optimized and GRASP/HP comes mainly from the coarse mode, and the AODF provided by GRASP/Optimized and GRASP/HP has almost no BIAS and show statistic that even better than AODF from GRASP/Models. This observation suggests that the approach of GRASP/Optimized and GRASP/HP can likely be improved in the future.

By the way, this result shows also the importance of keeping the section of fine and coarse mode validation for helping to improve understanding of the overall retrieval performance.

Line 592: But you compare against AERONET L2 which only includes AOD > 0.4. So, I don't understand the low AOD filter over ocean. because by comparing to AERONET L2 only large AOD cases will be included in the end. Or am I missing something?

Response:

Yes, we are comparing against AERONET L2 inversion data, 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).

Line 655: What do you mean by this?

Response:

We rewrote the sentence as follows. "MAIAC products cover some land-containing ocean tiles, however due to limited coverage of these retrievals we do not consider MAIAC ocean products here."

Line 717: It would be rather straightforward to compure GRASP AOD at 470 and 660 so that the exact same quantity can be compared.

Response:

In the paper, we intended to present the standard AE products validation for each dataset. Hence, we decided to use different wavelengths to compute AE for POLDER/GRASP and MODIS. Figure R1 shows the comparison of POLDER/GRASP AE (470/660) against with AERONET AE (470/660). This comparison is done by interpolating both POLDER and AERONET AOD to 470 and 660 nm based on the nearest available wavelengths. It presents similar performance with GRASP AE (440/870) in Figure 11.

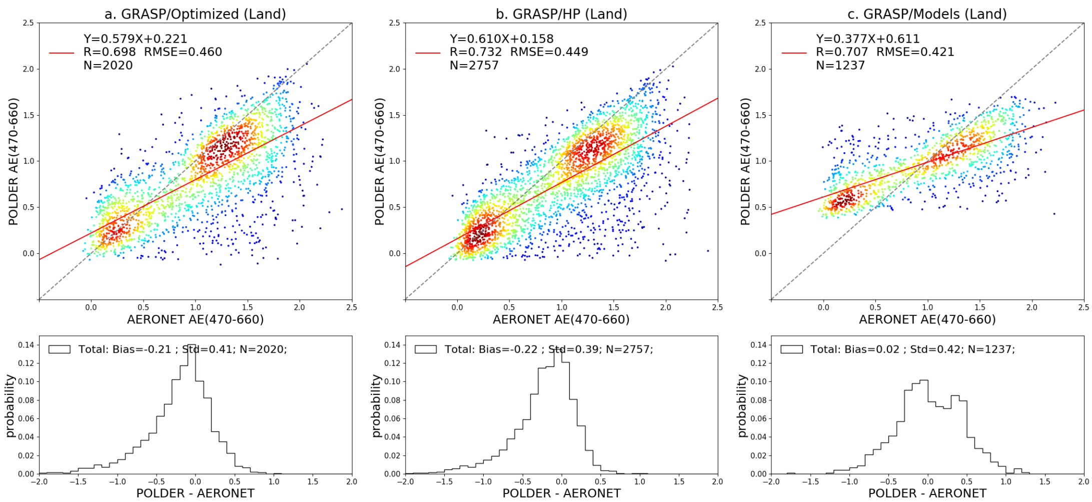

Figure R1. Validation of POLDER/GRASP (a. GRASP/Optimized, b. GRASP/HP and c. GRASP/Models) AE over land in 2008.

Line 899: decrease?

Response:

Thanks. It should be 'decrease'. It was revised.

Line 948: I would say the other statistical parameters are more important to discuss than the correlation.

Response:

In the revised manuscript, we have adjusted the discussion with including more comprehensive discussion of other parameters, e.g. R, RMSE, BIAS, GCOS, etc.

Line 1055: This needs more discussion. The bias between GRASP/Models and GRASP/HP is larger for most of the globe than between GRASP/HP and DT, right? This is quite unexpected given the AERONET comparison.

Response:

In Figure 24, indeed the differences between GRASP/Models and GRASP/HP are larger than DT and GRASP/HP. Table 18 confirms the difference (GRASP/Models - GRASP/HP) is 0.12 over land, and DT - GRASP/HP is 0.11; over ocean, the difference (GRASP/Models - GRASP/HP) is 0.35, and DT - GRASP/HP is 0.25. This is due to the phenomenon that GRASP/Models tend to overestimate AE for coarse particles (AE<~1.0) (see Figure 3). While in terms of correlation, GRASP/Models and GRASP/HP (Table 18) agree well (R>0.7) over both land and ocean, which can be interpreted for agreement in qualitatively indicating fine or coarse model aerosol. Overall, GRASP/Models can be improved by adjusting adopted aerosol models specifically for coarse mode models.

Line 1124: It is important to highlight the better AOD performance of GRASP/Models than the other GRASP versions before going to the comparison with other data sets. Namely, the better performance holds for GRASP/Models but not always for the other versions.

Response:

We now clarify it in the text as follow.

"*the PARASOL spectral products including AOD for six wavelengths in the range 443 to 1020 nm agree well with AERONET AOD measurements, e.g. for PARASOL/Models AOD correlation coefficients R are ≥ 0.86 over land and ≥ 0.94 over ocean with BIAS not exceeding 0.01 over land and 0.02 over ocean for all wavelengths. PARSOL/Optimized and PARASOL/HP also show good agreement with AERONET for spectral AOD, however they have non-negligible bias ~0.05-0.07 spectrally.*"