Reviewer #2:

Dear Kirk,

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

First and foremost, I'm not sure this manuscript is within the scope of ESSD. The "Aims and scope" portion of the ESSD website (https://www.earth-system-sciencedata.net/about/aims_and_scope.html) says: *"Earth System Science Data (ESSD) is an international, interdisciplinary journal for the publication of articles on original research data (sets), furthering the reuse of high quality data of benefit to Earth system sciences. The editors encourage submissions on original data or data collections which are of sufficient quality and have the potential to contribute to these aims. The journal maintains sections for regular length articles, brief communications (e.g. on additions to data sets) and commentaries, as well as review articles and special issues. Articles in the data section may pertain to the planning, instrumentation, and execution of experiments or collection of data. Any interpretation of data is outside the scope of regular articles. Articles on methods describe nontrivial statistical and other methods employed (e.g. to filter, normalize, or convert raw data to primary published data) as well as nontrivial instrumentation or operational methods. Any comparison to other methods is beyond the scope of regular articles."* I would think that an ESSD style manuscript on this topic would simply describe the various GRASP algorithms (briefly) and where they are archived. The majority of the paper is indeed comparisons to other data sets. I think it is far more appropriate for Atmospheric Measurement Techniques, for example (to pick another Copernicus Journal). I imagine this is something the editor needs to weigh in upon, but I wouldn't look to ESSD for this type of manuscript.

Response:

Multi-angular polarimetry (MAP) is always considered ideal for comprehensive retrieval of aerosol properties. GRASP algorithm was developed originally for operational processing of MAP measurements (Dubovik et al., 2011; 2014). The goal of this study is to announce the release of three archives of multi-angular polarimetry POLDER aerosol products processed by GRASP algorithm and provide comprehensive evaluation of these products against ground-based AERONET dataset, and popular MODIS aerosol products from DT, DB and MAIAC algorithms. For example, we found out that the quality of AOD retrieval from MAP (e.g. POLDER) is at least comparable to those of MODIS like imagers. In addition, we show that the MAP observations provide more information on detailed aerosol properties, e.g. spectral fine/coarse AOD, AE, as well as aerosol absorption properties such as AAOD and SSA. In this way, we assessed the potential of MAP sensors for aerosol monitoring. These both aspects are not surprising and were already discussed intensively in aerosol community. At the same, the absence of actual product from MAP sensors has often used as an argument for suggesting some overstatement of MAP potential. In these regards, our paper is aimed to answer this pessimism.

Several of our colleagues and co-authors suggested publishing our paper in new ESSD journal. Additionally, we were also inspired by your paper (Knobelspiesse et al., 2020) published in Earth System Science Data descripting the ACEPOL (Aerosol Characterization from Polarimeters and Lidar) field campaign, both of them show advances for aerosol characterization by utilizing the new era of MAP measurements from different perspectives. After additional consideration, we admit publishing our paper in other journals could be appropriate, but we remain convinced that this manuscript is rather appropriate for the Earth System Science Data. In addition, given the fact, that it was already exposed in open discussion and received several reviews, we prefer to continue with this journal.

Secondly, about the length. The manuscript review version is 108 pages long with 23 tables and 28 figures. You've chosen an extensive set of data to compare and contrast the various versions of POLDER data to the various versions of MODIS and then again to AERONET. What I'm missing is a concise set of objectives and how those are met. While I'm sure they are in the manuscript, they are lost among the noise. Because of the scale of the task, you need to be creative in finding ways to condense all of this analysis into something that easily and simply supports your work. Given the above two points, you may consider splitting this manuscript. For example, you could make an ESSD manuscript that describes basics of the dataset and its creation and archive location. It would point to one (or more) manuscripts that contain analysis. One could be on continuity with MODIS (and VIIRS?) and another on the full set of PARASOL products and comparison to AERONET. This is just a suggestion.

Response:

We agree that the manuscript is a long and that for the reader may be not easy to follow all details of the manuscript. In order to address that, we have revised the manuscript by combining all comments from reviewers, and trying to make it more readable. Generally, there are 3 main parts of this manuscript, (1) validation three archives (Optimized, HP and Models) PARASOL/GRASP products (spectral AOD, AE, AODF, AODC, AAOD and SSA) against AERONET data for entire PARASOL 2005-2013; (2) comparison of results obtained from validation of PARASOL (GRASP and Operational) and MODIS (DT, DB and MAIAC) aerosol products against AERONET in year 2008; (3) Inter-comparison satellite products at global pixel-to-pixel scale. The first part was on the full set of PARASOL products and comparison to AERONT. The second part tried to compare PARASOL and MODIS aerosol products by validating with AERONET follow the same criteria. The third part was inter-comparing PARASOL and MODIS aerosol products over globe at pixel level. We agree that they can be split in different papers but the separation of the materials of the paper is very difficult, because these three parts are quite complimentary. We feel that having the 3 parts together make the story more complete and after considerations we prefer to not split the paper.

Since the core of this manuscript is comparisons of other datasets, the choice of statistical metrics is very important. Table 3, for example, shows the use of nine different metrics, although Pearson's linear correlation coefficient is most commonly employed in the text. Unfortunately, some of these metrics, and especially the linear correlation coefficient, are not suitable for use on non-gaussian distributed data. The correlation coefficient is an expression of association, not agreement (Altman and Bland, 1983 and Bland and Altman, 1986), and is subject to numerical distribution, outliers, and sample range. So, for example, the R values for SSA are lower, but is that because of the lower success of the PARASOL retrieval (what you want to know) or because of the truncated numerical distribution of SSA which makes it non gaussian distributed? Additionally, what threshold of R can be considered a success? To that end, I think your metric for percentage within the GCOS requirements is a much more appropriate measure, and should instead be emphasized, although you need to take care to account for measurement uncertainty in both POLDER and MODIS or AERONET. Seegers et al. 2018 is a nice overview of these issues in ocean color data products, but equally appropriate here. They also identify regression slope and root mean square error as problematic, while noting that mean bias (which you use) and mean absolute error as appropriate. Bland and Altman suggest something similar, also recommending the pairwise mean bias and the "limits of agreement" which is similar to the mean absolute error. Variable measurement uncertainty can also be incorporated into these techniques (Knobelspiesse et al, 2019) which addresses the salient question "Do measurements agree to within stated uncertainties?" Ultimately, you should revise (and perhaps simplify) the metrics you use to assess your results.

Altman, D. G. and Bland, J. M.: Measurement in medicine: the analysis of method comparison studies, The statistician, 307–317 , 1983.

Bland, J. M. and Altman, D.: Statistical methods for assessing agreement between two methods of clinical measurement, The lancet, 327(8476), 307–310 , 1986.

Knobelspiesse, K., Tan, Q., Bruegge, C., Cairns, B., Chowdhary, J., van Diedenhoven, B., Diner, D., Ferrare, R., van Harten, G., Jovanovic, V., Ottaviani, M., Redemann, J., Seidel, F., and Sinclair, K.: Intercomparison of airborne multi-angle polarimeter observations from the Polarimeter Definition Experiment, Appl. Optics, 58(3), 650–669 , 2019.

Seegers, B. N., Stumpf, R. P., Schaeffer, B. A., Loftin, K. A., and Werdell, P. J.: Performance metrics for the assessment of satellite data products: an ocean color case study, Optics express, 26(6), 7404–7422 , 2018.

Response:

We fully agree that to choosing the adequate statistic metrics is very challenging. Therefore, for addressing this challenge we have presented many parameters at the same. In our understanding this approach allows us to have more comprehensive evaluation of the comparison results. Indeed, each single criterion has some limitations. For example, we agree that GCOS requirement is a good measure for AOD comparison, however, the total GCOS value tends to bias to small AOD, since >70% cases are coming from AOD<0.2. In Figures 7 and 8, GRASP/Models, MODIS DT, DB and MAIAC are all showing GCOS>45% for all AOD cases; while for AOD>0.2, most of them having GCOS<30%. The root mean square error (RMSE) also tends to be smaller for the products dominated by the results at lower AOD. The

BIAS may be misleading in cases when many deviations with opposite sign are added together. For example, it is often a case for AE. In Figure 3, GRASP/Models tend to overestimate of small AE and underestimate high AE, and total BIAS is close to ~zero, which is smaller than GRASP/Optimized and GRASP/HP. Another example, it the Table 10, where for very bright surfaces, many retrievals have positive bias that is compensated by the negative bias at higher AODs. Thus, in the revised version of the paper, we have tired not to focus the discussion on a single parameter throughout the manuscript. Also, we wanted to make sure to provide all parameters, such as correlation coefficients, that are traditionally used in satellite comparisons. This helps us compare our results with published ones. In addition, following your recommendations, we have revised the expression to emphasize on all evaluation metrics, e.g. R, RMSE, BIAS, Slope, Offset and GCOS.

Co-author Sayer has a publication about the numerical distribution of AOD and its impact on averages of data which is relevant to your matchup methodology and your use of Level 3 products. Were you calculating arithmetic or geometric means? I assume they are arithmetic since you don't mention otherwise, however this can in some cases cause an artificial bias in the results.
Sayer, A. M. and Knobelspiesse, K. D.: How should we aggregate data? Methods accounting for the numerical distributions, with an assessment of aerosol optical depth, Atmos. Chem. Phys., 19(23), 15023–15048, https://doi.org/10.5194/acp-19-15023-2019, 2019.

Response:

Yes, the L3 0.1 degree products were calculating using arithmetic mean and gdalwarp regridding technique (https://gdal.org/programs/gdalwarp.html). We have included this information in the text. As the spatial resolution is fine (compared to typical satellite composites at 1 degree), the arithmetic vs. geometric differences (as discussed in Sayer & Knobelspiesse, 2019) is likely significantly smaller for the present case.

I'm a little surprised that you make no mention of the retrieval algorithms for PARASOL (Hasekamp et al. 2011). In the beginning of section 2.1 you mention "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, and that it can perform it online without the use of traditional LUTs." You show that GRASP is an extremely power retrieval algorithm, but it is certainly not unique in its use of iterative RT computations, and I find it problematic that you make this claim and do not mention similar algorithms. Without acknowledging that there are other algorithms, even for POLDER, the manuscript sounds more like a sales pitch for GRASP and less an dispassionate piece of peer reviewed literature. In addition to Hasekamp et al, 2011, which is applied to POLDER/PARASOL, many others come to mind including those listed below.

Di Noia, A., Hasekamp, O. P., Wu, L., van Diedenhoven, B., Cairns, B., and Yorks, J. E.: Combined neural network/Phillips–Tikhonov approach to aerosol retrievals over land from the NASA Research Scanning Polarimeter, Atmos. Meas. Tech., 10(11), 4235–4252 , https://doi.org/10.5194/amt-10-4235-2017, 2017.

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 , https://doi.org/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, Optics express, 26(7), 8968–8989 , 2018.

Hasekamp, O. P. and Landgraf, J.: Retrieval of aerosol properties over land surfaces: capabilities of multiple-viewing-angle intensity and polarization measurements, Appl. Optics, 46(16), 3332–3344 , 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 , 2011.

Response:

We fully agree with this criticism and revised the text 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.