Atmos. Meas. Tech. Discuss., 6, C1163–C1167, 2013
www.atmos-meas-tech-discuss.net/6/C1163/2013/

Atmospheric Measurement Techniques Discussions

Interactive comment on "Aerosol retrieval experiments in the ESA Aerosol_cci project" *by* T. Holzer-Popp et al.

T. Holzer-Popp et al.

thomas.holzer-popp@dlr.de

Received and published: 1 June 2013

Jeff, we thank you for your helpful comments and hope that we can convince you that this paper as first of 3 steps led to significant improvements.

J. Reid (Referee #1)

Included here are selections from my "pre-review" which was submitted in the first stage of the review process. Being a pre-review, the authors were not bound to respond, and I see from the latest round by and large they hadn't responded to the major comments. Much of what I said then holds still today, so by and large I submit it again. Bottom line is that this is really just a report for the project, but without any solid substantial scientific results-one month of data is not enough to say concretely anything about products

especially with the coverage issues associated with AATSR . Alternating 6 months is the minimum before other than that from operational NASA products point of view, the ESA product lines have a long way to go. This is by no means a snarky comment, but a simple fact. I encourage the team members to apply lessons learned from the NASA development side, and in due time I think the ESA time series too will be valuable climate data sets. This paper is a report on the first round of intercomparisons and sensitivity tests of the ESA Aerosol_cci project. In summary, they examined September 2008 globally for a number of sensors and algorithms. For this one month, global plots are given for a) the native algorithms b) Algorithms run with identical aerosol microphysical models. c) A repeat of b) with a climatological aerosol prior to select the optical model, and d) a repeat of c with identical cloud masks (it is not entirely clear to be that it was not b)). From this the do brief comparisons and concluded that in general the retrievals improved by the use of the updated optical models, and it was a mixed bag for the use of the climatological prior. In the most general sense, this paper represents a progress report for the project. A three month followed by a 1 year examination is in the works. They do a nice job explaining the different algorithms and what they did. In so much that this is really a report, in itself I have no objections to it moving on to fuller review. This said, the science of this report, and in particular in regard to methodologies for verification and evaluation, is really quite poorly described.

First and foremost, while I understand how much work is required for a number of small shops to do a global analysis (the amount of data to be moved around is enormous), one month is not enough to say anything substantial with global conclusions. This is especially true with such narrow swath instruments such as used here. It is like trying to verify MISR with 1 month of data, something the MISR team would not do. What is presented is the grossest of "sniff tests" for the products. Since using improved optical models appears to help, this is not entirely wasted time, but if algorithms are to be 'judged' to determine which one will go into production, they are doing a disservice to the developers and sponsors alike.

On the analysis of only 1 month of global data: We are convinced that the 1 month of September global data cover many typical climatological and aerosol regimes and allow thus a first and fast assessment of the algorithm key elements (not a full-fledged validation and stringent selection of best algorithms, but a semi-quantitative comparison of different versions leading to better understanding of sensitivities). We saw in the further analysis that the results of validating the round robin algorithms with 4 months (not 3 – important that we analyse 1 month in each season!) and validating them with complete 12 months of the same year yielded pretty much the same statistical results. Knowing the results of the subsequent step (the round robin exercise with 4 months of data as described in de Leeuw et al, 2012 – paper accepted in RSE) and also the final step (analyzing 12 months of data, publication in preparation) we see that the analysis of the different algorithm versions in this paper has helped to identify possible improvements demonstrated in this paper plus additional necessary improvements identified in this paper such as post-processing for cloud contamination. In such the 1 month experiments constitute an essential step of a process which in the end led to 3 AATSR algorithms with quality equal to MODIS / MISR and much improved over the starting point of the baseline algorithm versions. For illustration of this final achievement we add here one result (global over land) from the publication in preparation analyzing the 3 AATSR algorithms and MODIS / MISR against AERONET.

Algorithm name / version Number of points correlation RMS Normalized mean bias AATSR_ADV.v1.42 1394 0,822 0,102 -29,7 AATSR_ORAC.v2.02 1394 0,823 0,091 -9,4 AATSR_SU_v4.0 1394 0,863 0,081 -7,7 MISR_V31_1x1 276 0,856 0,085 -11,2 MODIS5.1aqua 1185 0,749 0,114 7,1 MODIS5.1terra 1285 0,744 0,114 1,5

Based on the access review we had added a paragraph discussing the appropriateness of analyzing 1 month of data to the conclusions. We will carefully reword from “validation” to “sensitivity” to make clear the semi-quantitative nature of this analysis, but also stress that this assessment has indeed helped to understand sensitivities as initial step of the 3-step process, which ultimately led to significant algorithm improve-

ments proven with 12 months of data.

We will add some more detail on the validation methods (the focus of this paper is not to develop new validation techniques)

Specific issues for consideration include 1) Based on Figure 5, it seems that there is very little data making it into the composite plots. Contextual bias is likely severe. A map of how many samples actually go into the map is certainly required. I imagine it is on the order of 3 or 4 over the non-arid parts of the world. Is this really enough to say anything concrete? The purpose of showing the maps was to illustrate how well an algorithm version can reproduce the major spatial distribution patterns. This aspect is not well covered by AERONET stations statistical analysis with large spatial gaps. We will add maps of numbers. Note that the AERONET validation was done on basis of daily data.

2) For AERONET, the scientists here have fallen into the same trap as the MODIS folks. Namely, MODIS optimized their retrieval to the global data set. But, the overwhelming majority of sites is in the US and Europe. Thus, when regional studies were conducted, the algorithm failed everywhere except eastern US and Europe. You can show an improvement in bulk scores by using their optical models, but in reality make things worse in certain locations. Certainly, they need to break things down to areas of a dominant species (smoke, biomass burning, etc). Also for AERONET, they are using log scales. Their RMSE is large enough that they can plot this linealry. As we do (look at Yingxi shi’s papers) I would do a linear plot, with a regression line and RMSD bounds. We will add analysis broken down to land and coastal stations as well as several regions outside Europe and North America and to some extent representative for different dominant species to prove that the optimization was not only done towards European and North American aerosol regimes. It should be noted that the common aerosol models used cover all major aerosol types globally (also outside Europe and North America). There are only few AERONET ocean sites (MAN data in 2008 are not sufficient in coverage), so that only a differentiation between coastal and inland sites

will be possible. Scatter plots will be replaced by those using linear scales.

3) Over both land and water, it is clear from the plots that the lower boundary condition algorithm components are deeply flawed for all algorithms. Over land this is understandable. Over water, this is about as slow of a pitch possible. Even more discussion of the lower boundary condition needs to be included, as it is what is driving verification bus. The second and third last paragraph of the conclusions discuss exactly this very important aspect over land and ocean and why it was not yet assessed in this paper. Seeing the results of de Leeuw et al., 2012 (where the AATSR algorithms developed based on the preparatory analysis of this paper and further improvements deduced from this paper reach the accuracy level of MODIS and MISR) we disagree with the statement that the lower boundary condition is deeply flawed for all algorithms; this statement is true for the nadir only (MERIS, SYNAER) algorithms as it is stated in the conclusions (third last paragraph) but not for the dual or multi view sensors AATSR and POLDER.

4) In general I found the figures small and very hard to read. Some effort needs to be made in enlarging and improving figure quality. I see they took the comment that they should use one color bar, but now the figures are a bit awkward. My suggestion (To one and all in our community) is the money spent on Adobe Illustrator is money well spent. In addition to colour bars changes, we have already enlarged and re-arranged figures 6-14. However, we will make another effort in improving their

---

Interactive comment on Atmos. Meas. Tech. Discuss., 6, 2353, 2013.