

Interactive comment on “Inter-comparison of three AATSR Level 2 (L2) AOD products over China” by Y. Che et al.

Anonymous Referee #2

Received and published: 25 March 2016

This paper validates 3 algorithms (SU, ADV, ORAC) for determining AOD from the European AATSR sensor against Sun photometer data in China (from AERONET and CARSNET). The topic is relevant to ACP. The work is important because these European products have not been as well known as NASA ones, and have undergone a lot of development in the European CCI projects, so it is a good time to do some more thorough validation of these data sets. This is especially true for China since the aerosol loading is high and variable, and CARSNET has monitoring stations in some areas where AERONET is lacking.

I have read through the manuscript several times and, while it is promising, there are some things which are unclear/invalid or I think not useful, and some important things which should be added to make the analysis more complete and useful. The phrasing is odd in some places and there are a number of typos (e.g. AEROENT in some places

C1

instead of AERONET) so I think that the manuscript will need some copy editing by the production office. I appreciate that English is not the authors' first language and the writing is not bad, it is just unclear in some cases. I therefore recommend some content revisions, to address the points below. I would like to review a revised version and think that another round of reviews will be necessary because the structure/content of the manuscript might change a lot. Here are my main points:

1. Abstract: some of the sentences could probably be shortened (e.g. first and second can be combined, as can third and fourth).

2. Introduction and start of section 2: It would be good to add a bit more information about the AATSR sensor here, like launch/end dates, swath width. A brief discussion of the differences between the algorithms should be included, to help understand why they give different results. From the analysis, the performance and the spatial coverage are both different between the three algorithms, so some insight into what in the algorithms is responsible would be welcome.

3. Statistics. Some of the metrics presented here are questionable in their relevance and I think that there are simpler and clearer alternatives. Specifically, the EE envelopes quoted here for Equations 1, 2 are for the MODIS instrument, not AATSR. AATSR is quite different (two views, fewer wavelengths) so there is no reason to expect that an AOD retrieval for AATSR would have the same type of behaviour. One might expect that the error formulation would be closer to that of MISR. Further, Equations 3 and 4 are basically expressing a confidence envelope around a regression line. This is not really useful since it is just the noise around the relationship and not so dependent on the actual error in the retrievals. So comparing this between algorithms does not really make sense. A well-correlated but very biased retrieval would appear 'better' by this metric than a poorer-correlated but unbiased one, while for an actual scientific application, the unbiased one may in some cases be more useful.

Further, least-squares linear regression is invalid for AOD retrievals be-

C2

cause aerosol data violate the assumptions of this technique (see e.g. <http://people.duke.edu/~rnau/testing.htm> ; the AOD data validate assumptions 3 and 4 that linear least-squares regression makes, possibly 1 and 2 as well, and as a result the results obtained are not statistically valid). I know that a lot of people do least-squares linear regression because it is easy, but it is still wrong for this application.

So, a better alternative is just to present statistics of bias and RMS error as a function of AOD, similar to what is shown in e.g. Figure 5 and the magenta bins in Figure 2. So I suggest that the EE discussion here and linear regressions be discarded entirely, and more prominence should be given to statistics subset into different regimes (e.g. low AOD, moderate AOD, high AOD; perhaps also splits based on Angstrom exponent for the high-AOD regime, as retrieval errors are often type-dependent as well). The kappa coefficient is probably fine. So, accounting for this comment would somewhat streamline and improve sections 2 and 3.

Finally, presenting statistics to 3 significant figures is overkill and paints a picture of them being more robust than they probably are; 2 significant figures is probably enough.

4. Retrieval errors. As I understand it, the CCI project means that the data products also provide their own estimates of the uncertainty in the AOD retrieval for every pixel. This is an important point, since pixel-level uncertainties are very useful for many applications. However, it is not discussed in the paper. How do these uncertainty estimates compare to the retrieval errors observed?

5. Figures 2-4 and discussion in Section 3. I don't see any advantage to splitting out the points by year. It would be easier to combine all points from one algorithm into one panel, not 3. This would also let you combine Figures 2-4 into one figure for a side-by-side comparison of the three algorithms. Also, as discussed before, I would delete the regression and EE lines here since they are not very meaningful. The magenta symbols and lines for the binned data are enough here. Also, the color scales used in

C3

these figures are not mentioned and can probably be removed (either show individual points without a color scale or a density plot with a color scale).

I also don't see any good reason to split the discussion of statistics out by year either. The data volume is not very large, so year to year differences are probably resulting from sampling and not statistically meaningful. Looking at the bigger picture of all data together is more statistically robust and gives a clearer picture. I don't believe any insight is gained by splitting the analysis up year by year.

6. Figures 5-7: Similar to the last comment: are the different panels the different years? It doesn't say anywhere but I infer that is the case. Again, these figures could be streamlined into one because clearly the biases are similar between years, this will be more robust, and will allow for a more direct comparison of the 3 algorithms. Additionally, I don't think the histograms (bottom panels) here are useful since they don't provide any information which is not seen clearly in the top panels, so these could be deleted. Also, for the same reason as before, the linear regressions are invalid and should be deleted, just showing the binned values is enough.

7. A similar plot to the bias plots could be created for RMSE. This would be a clearer way to show and compare the AOD-dependence of the retrieval error than the EE3/EE4 metrics.

8. In the discussion of the results, a lot of the time terms like "good" and "well" are used to describe performance. These are "weasel words" and should be avoided. What is "good" is only really relevant relative to a specific application (e.g. good enough to do X) or compared to the state of the art. I suggest rewording to avoid these words and be more quantitative where possible, or else stick to comparative terms (e.g. say when the data sets are similar to or better than each other). Also, some discussion of results compared to validation of other sensors (the main ones being MODIS/MISR) could be included, as these all have published validation for their aerosol products, and this would give a sense of how the AATSR data perform relative to the other available

C4

data products. Right now the paper more or less reads like AATSR is the only satellite option.

9. Do the retrievals provide other information like Angstrom exponent? From other references, I believe so. This quantity is commonly compared with AERONET measurements, so it should be easy to extend the analysis to look at this as well using the same basic approaches. This might provide more insight for the differences between the data sets, if the algorithms are making very different assumptions about what sort of aerosol is present. This would help overcome one of the weaknesses of the paper, i.e. that the comparison is presented without any sort of discussion about why the three data sets are different and how to improve them (which would be very useful information).

10. There are at least two more AERONET sites in China which provided data in the study period, but which were not used in the analysis. These are both in Hong Kong: Hong_Kong_Hok_Tsui and Hong_Kong_PolyU. Why were these not used? If the objective (as stated) is to provide coverage over broad areas of China, then it would make sense to include them, since the data are freely available and there are no other sites used in this part of China. These sites are very close to the coast so also provide an additional type of environment to analyse, compared to the other sites presently included in the study. Additionally, it will boost the data volume. I suggest adding these sites to the analysis. There may be more, these were the main ones which sprang to mind. On a related note, Figure 1 can probably be simplified for clarity by using one symbol/color for all AERONET sites, and another for all CARSNET sites. Splitting by year isn't necessary, in my view, and just complicates things.

11. The title of the manuscript suggests a broader scope than the analysis, since the analysis only performs an inter-comparison in the context of AERONET/CARSNET measurements. There are various other things which could be added, at least briefly. For example, climatologies of seasonal AOD from all three algorithms (from the 1 degree products), and maps showing the available data volume (e.g. number of days per

C5

season with data), since this is another feature which is important for many applications. Otherwise, the title should be amended to reflect the scope. However I would prefer that the analysis be extended because I think that this would be quite useful (and new, to my knowledge).

Interactive comment on Atmos. Chem. Phys. Discuss., doi:10.5194/acp-2016-195, 2016.

C6