Atmos. Chem. Phys. Discuss.,
doi:10.5194/acp-2016-748-SC1, 2016

Atmospheric Chemistry and Physics Discussions

Interactive comment on "Assessment of the aerosol optical depths measured by satellite-based passive remote sensors in the Alberta oil sands region" *by* Christopher E. Sioris et al.

A. M. Sayer

andrew.sayer@nasa.gov

Received and published: 14 October 2016

I read this paper with interest, as I am involved with the development of the Deep Blue aerosol products and was recently a co-author on an analysis in this region (Li et al 2016), which the authors have also cited. I had some general comments and suggestions about the use of the satellite aerosol data products used, mostly MODIS. This should not be taken as a full peer-review (I'm not commenting on the PM parts). This is just a comment on the satellite data and AOD analysis.

I am glad to see that the authors used both the Dark Target and Deep Blue products

when conducting their study. However, I see they use the Dark Target AOD product at 470 nm, rather than 550 nm. 550 nm is the main reference wavelength for this product, the one that has been validated, and the one which is generally recommended to be used (and is indeed used by most data users). Keeping everything consistent at 550 nm (or a close wavelength, e.g. MISR's 558 nm) where possible also makes it a clearer comparison between the various products. My suggestion would be to do the analysis with the standard 550 nm AOD product. The Dark Target 550 nm AOD is also contained in the same multidimensional SDS that the authors used (Corrected_Optical_Depth_Land), or the authors may go directly to the combined Dark Target land and ocean SDS Optical_Depth_Land_And_Ocean, which contains AOD at 550 nm with the quality flags already applied (plus includes over-water retrievals). So no additional data download should be necessary to do this. I don't see any reason to actively choose the 470 nm AOD over the standard 550 nm for this analysis.

Similarly, the Deep Blue AOD quality flag is in Deep_Blue_Aerosol_Optical_Depth_550_Land_QA_Flag, but we also provide a data set which already has the quality flag mask applied (Deep_Blue_Aerosol_Optical_Depth_550_Land_Best_Estimate) so the user does not have to do the filtering themselves. It is not clear to me from the paper which SDS was used to QA-filter the Deep Blue data but I am assuming it is the above. More information can be found in the MODIS aerosol file spec document (http://modis-atmos.gsfc.nasa.gov/_specs_c6/MOD04_L2_CDL_2013_03_21.txt) or on our website, http://deepblue.gsfc.nasa.gov. Could this be clarified?

In terms of Aqua vs. Terra, Aqua has indeed been historically more stable and better-characterised so I agree that it is probably sufficient to use only Aqua for the analysis. Note that for Deep Blue we apply additional calibration updates to Terra which are not included by the Dark Target team in the present version of the processing (see Sayer et al JGR, 2015, doi:10.1002/2015JD023878), so our Terra/Aqua differences are smaller than they see and we do not have the same divergence in time through

most of the mission (although in the past year or so there appear to be troubles with Terra calibration again, which are under investigation).

Also, which ATSR product is used? There are at least 3 being produced in Europe in the framework of the ESA CCI project, and they all have different approaches and results (see Popp et al, Remote Sensing, 2016, doi:10.3390/rs8050421 for an overview). My inference is that this is the Swansea algorithm (Peter North's group) but I think this should be stated more clearly. Perhaps the others could be added to the analysis as well, if this is not too much effort. Similar to Dark Target vs. Deep Blue for MODIS, the various ATSR algorithms have different coverage.

For POLDER, the data product the authors have used reports AOD at 865 nm. Due to the wavelength dependence of AOD, in most cases this means that the AOD will be much lower at 865 nm than 550 nm. The smaller signal will probably cause problems for relationships constructed using this AOD, plus one would not expect a close match between AOD at 550 nm (given by the other sensors) and 865 nm since the spectral dependence of AOD is determined by the aerosol composition. I wonder if another POLDER data product like GRASP (see e.g. http://www.grasp-open.com/products/) which does report AOD at 550 nm would be more useful here (and also allow for a more direct comparison between the various data sets).

I had also been under the impression that the particular POLDER AOD retrieval data set the authors are using is intended to be only a fine-mode AOD retrieval, rather than a total-AOD retrieval, which further complicates things. However, I may be mistaken about that as I have not used POLDER data myself for a few years now.

I note in the text that AERONET AOD was interpolated to the satellite wavelengths (which is the standard practice), but Table 4's caption says that AERONET data at 500 nm were used. I guess that this is an error in the caption, but can this be clarified?

Figure 1: My guess is that the white spots on the maps for POLDER and MISR are because the level 2 retrievals are at a coarser resolution than the 0.1 degree grid being

averaged to. In that case it might be better to allow the level 2 data to occupy multiple grid cells (corresponding to the actual retrieval footprint) than to snap them to the grid cell nearest to the pixel centre (which is what I assume is being done here). If the retrieval pixels are larger than the grid size (which is the case here) then it does not really make sense to assign a pixel to one grid cell, when it occupies multiple grid cells.

As a general comment on this figure, I would recommend keeping the colour scales the same (and ideally start at zero) to allow a direct comparison between the different data products. Right now it is hard to compare them because the colour bars are different. I realise POLDER is the odd one out here since it is at a longer wavelength, but the other data sets (at or near 550 nm) should be on a consistent scale. I'd also suggest mentioning again in the caption that POLDER is at 865 nm, hence the lower AODs.

As another general comment on the above figure: we know there is seasonal variation in AOD, as well as variation in things that affect sampling (e.g. cloud and snow cover). So presenting an annual mean here conflates these issues together with the issue of retrieval uncertainty. My suggestion would be to make separate maps for each season. They don't all necessarily need to be included in the paper if length is a concern. This way the seasonal aspect at least can be removed and it may bring the different data sets into closer agreement (or it might not). The next stage would be to compare the points only where they have common retrievals on the same days, but I suspect that due to the large number of data sets there would probably be few mutual points. So, making seasonal means rather than annual means is probably a good balance in terms of seeing how the data look compared to each other.

Figure 2: If I understand correctly, this is the mean of the MODIS Deep Blue and Dark Target QA values. I understand the intent behind this figure (illustrate where the algorithms have confidence) but I think the execution is problematic. By taking the mean of the QA flag, it is being treated as a quantitative variable. However it is not – it is a categorical variable that is stored as an integer because it is easy to store integers in the hdf files. QA=0 has a fundamentally different meaning (no retrieval)

from the other values, and the QA from 1 to 3 does not represent linear progression in terms of quantitative retrieval quality or uncertainty. So, taking the mean value is a bit misleading since it is conflating lack of retrievals (due to e.g. clouds) with other algorithm factors and giving a number as a mean for the grid cell which doesn't really relate to the underlying QA flags. For example if the mean QA calculated in this way is 1, it does not mean that the retrievals here have low confidence. It means either that the retrievals have low confidence, or that there is some combination of high confidence retrievals and data gaps due to clouds, etc.

So, I think this figure should be updated, and we might get some more insight into what is going on if the metric here is calculated differently. In Deep Blue we recommend QA=2 and QA=3 can both be used for quantitative analyses as they have similar error characteristics (Sayer et al., JGR 2013, doi: 10.1002/jgrd.50600) while for Dark Target land retrievals they recommend QA=3 only (e.g. Levy et al, ACP 2010, doi:10.5194/acp-10-10399-2010). This is another example of the fact that QA flags have different specific meanings for different data products. What I would suggest is making maps showing the fraction of overpasses where there is no retrieval (i.e. QA=0), the fraction where there is a poor-QA retrieval (i.e. QA=1 for Deep Blue, QA=1 or 2 for Dark Target), and the fraction where there is a good-QA retrieval (i.e. QA=2 or 3 for Deep Blue, QA=3 for Dark Target). Those maps would reveal those areas where the retrievals frequently fail or are absent in a more meaningful way than the current Figure 2, in my view.

Some of the data holes in the MODIS Dark Target product will be from the fact that neither their land nor ocean algorithms treat pixels which are identified as 'coastal' as valid for AOD retrieval. (Note that Deep Blue treats such pixels as land, but excludes pixels next to water frequently for other reasons.) This limits coverage in many parts of Canada and elsewhere in the world, as pixels containing lake shores are frequently identified as coastal. See Carroll et al. (IJDE, 2016, doi: 10.1080/17538947.2016.1232756).

Figure 3: This shows that in areas where there are few AATSR retrievals, those retrievals that are performed tend to have a higher sub-pixel cloud fraction. The implication is that sampling in this area is influenced by cloud cover, whether real cloud or misidentified cloud (which is reasonable). However what might make a better right panel would be the cloud fraction for ALL observations, not just for those observations where an AOD retrieval is performed. This would look more directly at where the AATSR algorithm thinks there is a cloud. Right now what the panel is showing is subtly different since pixels which are cloudy above the threshold for retrieval (I am not sure if this is 100% cloudy or some lower fraction) are exclude from the analysis.

Table 1: Again, the MODIS standard AOD wavelengths for both Deep Blue and Dark Target are 550 nm. Deep Blue also provides 412, 470, and 650 nm and Dark Target also provides 470 and 650 nm. Source radiances are not all at 0.5 km pixel sizes, it depends on band, so it would be better to say 0.25-1 km here. Also, due to its scan design and wide swath with, MODIS level 1 and level 2 pixel size and shape get heavily distorted from nadir to scan edge (quoted values are all for nadir pixels), which is not an issue for AATSR or MISR to the same degree due to their designs and narrower swaths. See e.g. Sayer et al (AMT, 2015, doi:10.5194/amt-8-5277-2015) for more information.

Table 4 and discussion: I would delete the analysis of linear least-squares regressions from the table and discussion. AOD data violate most/all the assumptions required for this technique to be valid, and so the results are misleading and fits/confidence envelopes are quantitatively incorrect. See e.g. http://people.duke.edu/∼rnau/testing.htm for more discussion. (I know it is a frequently-used technique in our community, but it is fundamentally incorrect for this particular application.)

I hope these comments are useful; please feel free to get in touch if you have questions about them, or about the MODIS aerosol products in general.

Interactive comment on Atmos. Chem. Phys. Discuss., doi:10.5194/acp-2016-748, 2016.