Response to short comment by Dr. Andrew Sayer

We thank Dr. Sayer for his suggestions (in red).

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).

We agree with this comment. It was realized after the original submission of the manuscript that the 470 nm product was selected unintentionally instead of the 550 nm product. Rather than withdraw the manuscript or ask for a long extension to regenerate a decade of MATLAB *.mat files required as input for our validation and mapping software, the Dark Target AOD at 470 nm was retained temporarily with the full intention of redoing the map in Fig. 1, the validation results (Table 4), *et cetera*, at the next stage in the review process.

We now write at p2L28:

Specifically, the Corrected_Optical_Depth_Land (550 nm) and the Deep_Blue_Aerosol_Optical_Depth_550_Land datasets were used and confidence for both datasets was extracted from the Quality_Assurance_Land dataset.

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://modisatmos.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?

We agree that the ACPD manuscript fails to name the SDS used to QA-filter the Deep Blue data. 'Quality_Assurance_Land' is the SDS used.

The change to the manuscript is contained in the sentence mentioned above at p2L28, in response to the previous comment.

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.

The selected ATSR product is stated clearly in the appendix of the existing manuscript (p13L17) and the appendix is referenced at p2L26 in connection with the satellite data products. Information on the ATSR product is in the following sentence of the appendix (p13L17)

AATSR and ATSR-2 version 4.1 data are from Swansea University and can be obtained from the Aerosol CCI website (http://www.esa-aerosol-cci.org/) following registration.

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.

There are three different algorithms for both AATSR and ATSR-2, and at least two POLDER algorithms, several MODIS products (Terra vs. Aqua, Deep Blue vs. Dark Target), plus MISR. That is eleven, and it is not an exhaustive list of available products from these satellite-based sensors. The primary focus of this paper is not on algorithms but on the different aerosol sensors. The Swansea University algorithm was chosen since initially they had, by far, the longest AATSR data record available.
To make this decision clear, we now write at p3L9:

The focus in this paper is primarily on the different aerosol sensors, rather than the different retrieval algorithms applied to the same satellite data (e.g. Popp et al., 2016), with the exception of the widely used Deep Blue and Dark Target algorithms for MODIS.

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).

We have used the only POLDER AOD data product that was available at CNES's POLDER website. We did not search the web or the literature for alternate POLDER products.

The different satellite AOD data sets are essentially not compared in a quantitative way. The quantitative comparison is essentially against AERONET and thus the different wavelength (865 versus 550 nm) is not a major issue since AERONET measures at 870 nm and many wavelengths in the visible. The smaller aerosol signal at 865 nm does not cause problems for the linear regression relationship constructed between POLDER and AERONET AODs. This is obvious from the high correlation coefficients for POLDER in Tables 3-5. Also POLDER reports AOD at 865 nm, but uses measurements at 670 nm in the AOD retrieval.

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.

Dr. Sayer makes an interesting point here. This is not a fine-mode AOD product; total AOD is retrieved and reported. See:
http://www.icare.univ-lille1.fr//projects_data/parasol/docs/Parasol_Level-2_format.pdf.
However, the use of polarized radiances in the POLDER retrieval greatly reduces the sensitivity of the retrieval to coarse particles. Thus, it is possible that a coarse-mode aerosol plume could, to some extent, mask the polarization signal from underlying fine-mode particles if such an arrangement occurred. Ultimately, the low sensitivity of POLDER to coarse-mode particles appears to be a minor issue at the two AERONET sites (Fort McMurray and Fort McKay) given the lack of bias and the high degree of correlation with AERONET AOD, in spite of the fact that coarse-mode dust is known to be significant contributor in this region, particularly at Fort McKay (based on AATSR dust fraction, not shown).

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?

Dr. Sayer is correct that this needs clarification, even though there is not an error. AERONET 500 nm AOD is used, however it is scaled to the satellite wavelengths.
In the caption, we now write:

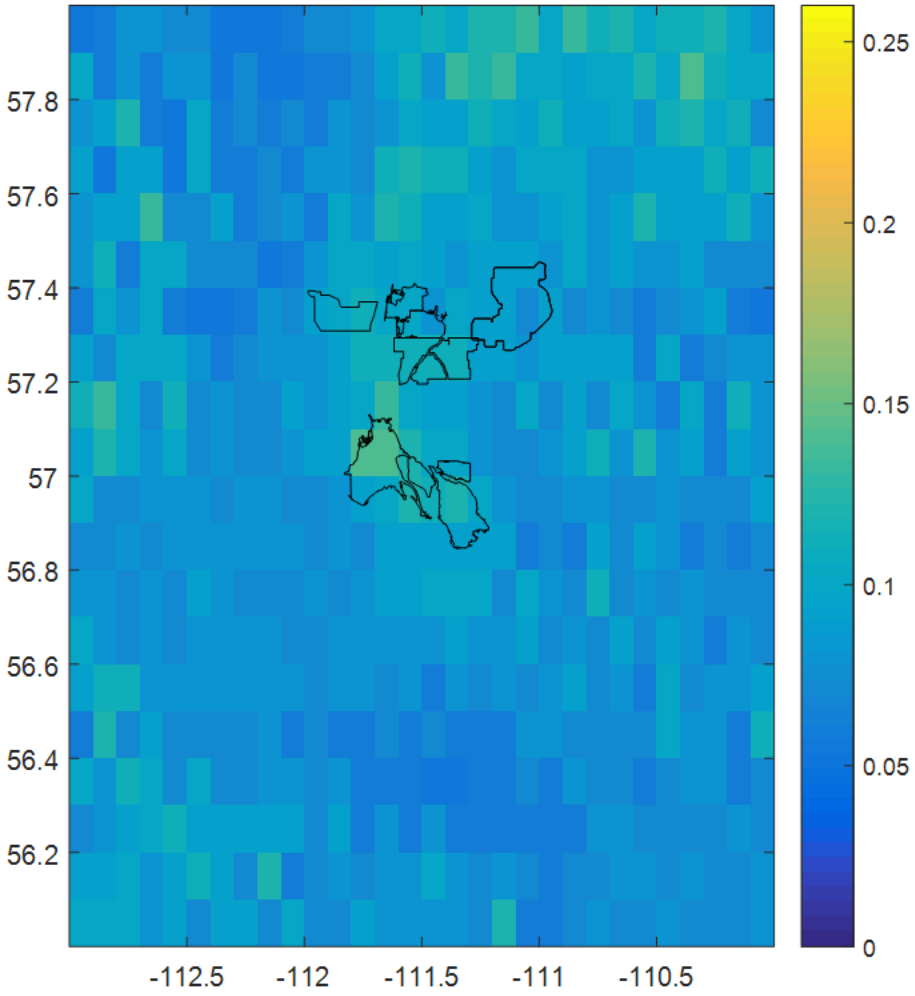
The Cimel 500 nm AOD, scaled to the satellite AOD wavelength (see Sect. 2), is used for comparison with all satellite sensors except POLDER/PARASOL, for which the Cimel 870 nm AOD is more appropriate (see Table 1).

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.

The orientation of actual footprint would need to be known and, for POLDER, this information is not available for each observation: only the latitude and longitude at the center of the AOD superpixel is provided. In general, we disagree that it does not make sense to assign a pixel to one grid cell. This is referred to as spatial oversampling and can be very revealing about localized sources of aerosols.

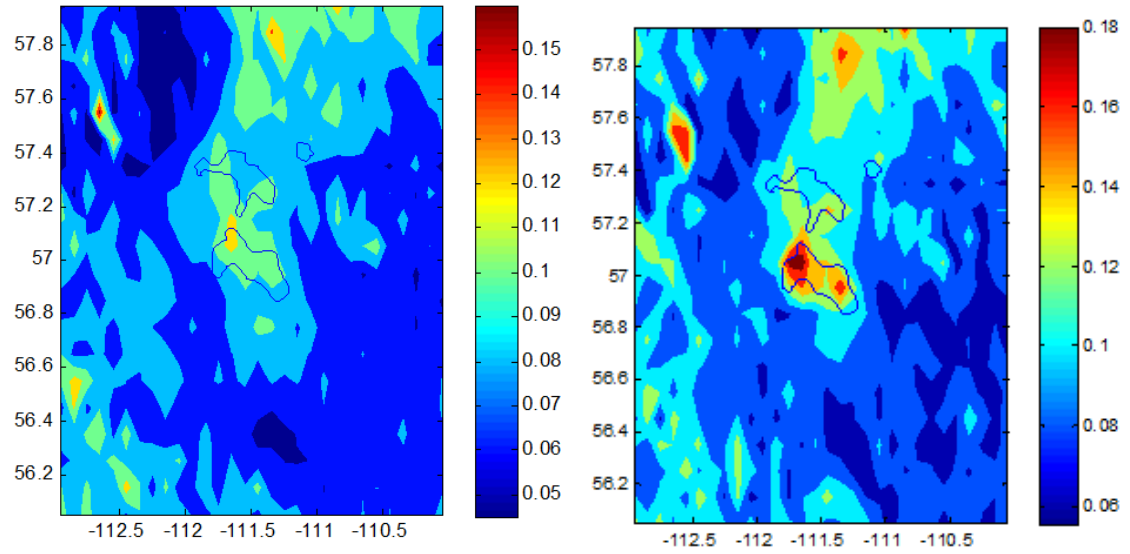
As a general comment on this figure, I would recommend keeping the colour scales the same (and ideally start at zero) to allow a direct comparison between the different data products. Right now it is hard to compare them because the colour bars are different. I realise POLDER is the odd one out here 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.

This recommendation initially seemed like a good one, but even the AOD differences between the MODIS products using the respective confidence values suggested by Dr. Sayer near the Syncrude facility are quite large, as shown in this Deep Blue climatological mean AOD map using confidence 2-3, but with the AOD range of the colour bar extending to 0.26 to cover the maximum climatological AOD of Dark Target (confidence=3).



Including such a figure would severely compromise our primary goal for Fig. 1, which is to show the spatial gradients in AOD in this region. The colour scales have been changed to have a common lower limit of 0.

We already mentioned in the caption that POLDER is at 865 nm: "(top left) POLDER 865 nm (1996-2013)". Just as a point of information, the Deep Blue climatological AOD for confidence=3 has a hotspot near the Suncrude facility with AOD of 0.12, yet we find that higher climatological maximum AODs occur (0.18) when only confidences of 1-2 are retained, again with the hotspot being the Syncrude Mildred Lake facility, as shown in the following maps to the left and right, respectively.



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.

We tried plotting AODs for May through September for the MODIS and MISR products (Figures A-C below). These are the months when all five aerosol products have high measurement frequency. But again, Dr. Sayer's purpose is evidently different than ours: we are not trying to bring the different data sets into closer agreement; as stated up front (p4L33), we are mostly trying to see what each is capturing spatially over the long term, so annual means are preferable. Anyway, as shown in Figures A-C below, limiting to these 'warm season' months does not bring the data sets into closer agreement. Limiting to the warm season was mostly expected to benefit the MISR AOD map since MISR has an unusual spatiotemporal sampling pattern, but as shown in Figure C, limiting to May-September does not produce a more coherent AOD map. In the revised manuscript, all available months are retained for Figure 1.

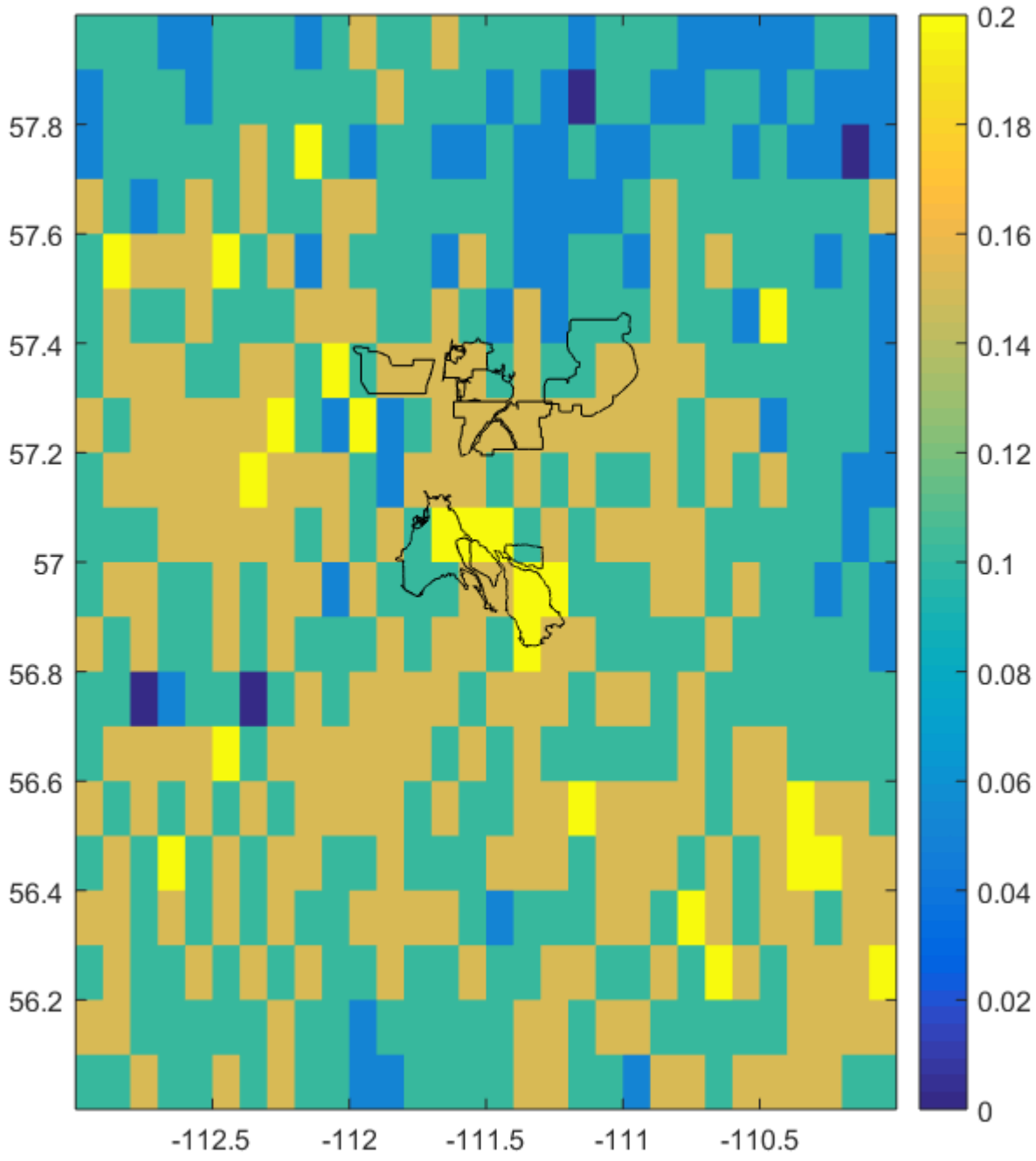


Figure A: MODIS DT 550 nm climatological AOD for May to September (confidence=3).

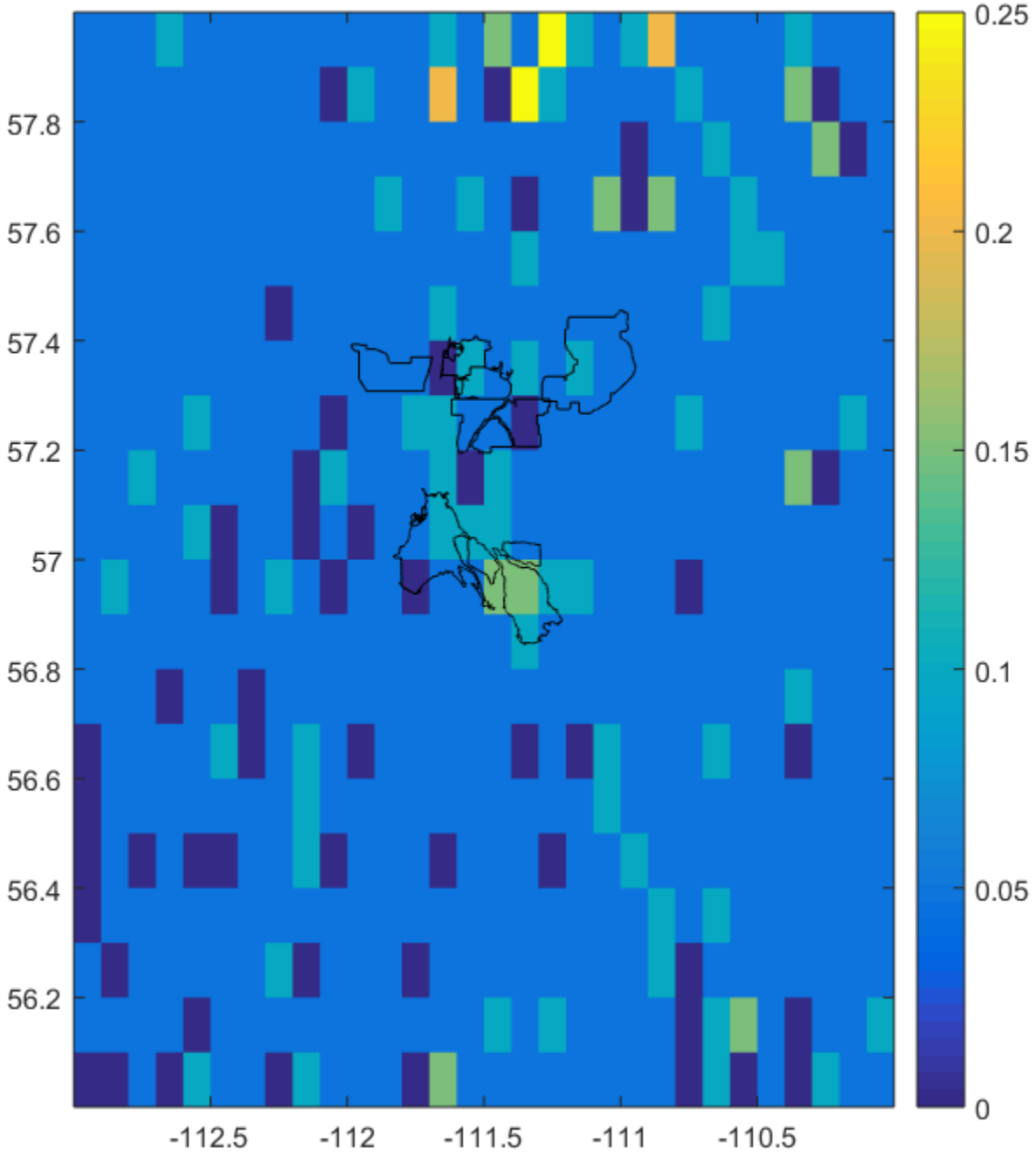


Figure B: MODIS DB 550 nm climatological AOD for May to September for confidence≥2.

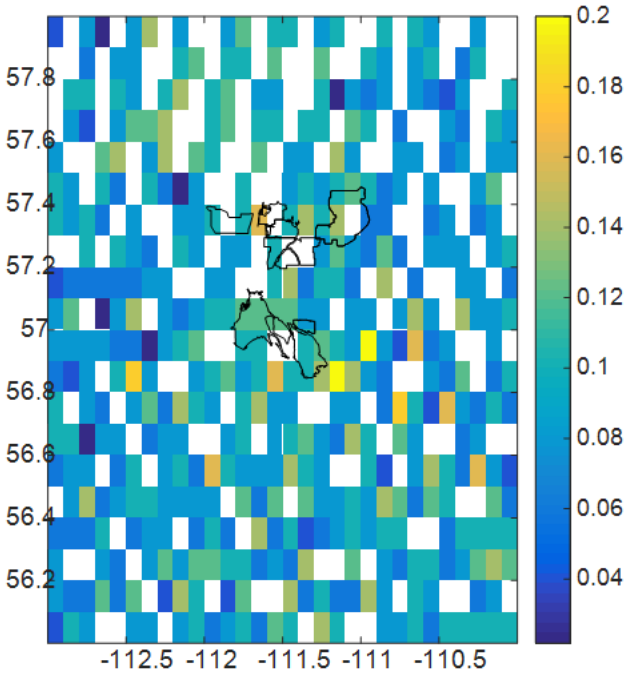


Figure C: MISR 558 nm climatological AOD for May to September.

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.

This comment by Dr. Sayer is correct, and we were aware of all of these logical points. The main purpose of both panels of Fig. 2 was to show that QA is tending very close to 0 (i.e. <0.45) at the two grid cells near the Syncrude Mildred Lake facility, implying that the retrieval has no confidence (or provides a fill value) more than 55% of the time.

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).

This suggestion is accepted. A new six-panel Fig. 2 has been generated.

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).

This cause of data holes has been added to the list of causes. We now write at p5L16:

The number of pixels used in the AOD retrieval is reduced by the inland water mask (Carroll et al., 2016), …

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.

Additional cloud tests (Bevan et al., 2012 and reference therein) were used for this AATSR aerosol retrieval algorithm that are not used in AATSR Instrument Processing Facility (IPF) v6.01 cloud product. Thus, we feel it is more appropriate to look at the cloud fractions in the successful AOD retrievals. This suggestion might have been worth pursuing if the spatial anti-correlation was not strong between cloud fraction in successful AOD retrievals and AOD sample size, but that is not the case.

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.

In Table 1, regarding the spatial resolution of MODIS radiances, we now write: 0.25×0.25 to 1×1. We have also changed one column heading to: "Spatial resolution of AOD superpixel at nadir".

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.)

Dr. Sayer's most recent paper (Carroll et al., 2016) cites Levy et al. (2013) for AOD validation, and Dr. Sayer is also a co-author in the latter work. This latter work includes linear least-squares regression of MODIS AOD and AERONET AOD (their Fig. 11), which is precisely what we have done. It is clear that our Table 4 adheres to the established convention in this field in terms of validation statistics. As an alternative, we tested two non-parametric methods (Theil's complete and incomplete methods) to obtain the values in the first three columns of values in Table 4. None of the assumptions are violated when using Theil's incomplete method (1950). Also, application of Spearman's rank correlation is valid for this application (see Table 4). As shown in the table below, the non-parametric methods yielded slopes that were small (~0.6) and ordinary least-squares ('OLS') yielded a slope that was clearly of the wrong sign due to one small cluster of outliers at high AOD. We tested a number of robust regression methods compared in Holland and Welsch (1977), which all use a weighted least-squares (WLS) approach to reduce the sensitivity to anomalous data pairs (i.e. coincidences). Some of these robust regression methods are expected to perform better than OLS on data with non-Gaussian distributions (e.g. Andrews, 1974). The outliers affect whether the AERONET and satellite AOD data conform to a normal distribution. The table below presents the slope and offset from various robust methods using the POLDER/PARASOL and AERONET coincident data at Fort McMurray:

Method	**offset**	**slope**
Andrews	-0.017	0.787
bisquare	-0.017	0.788
Cauchy	-0.017	0.797
Fair	-0.019	0.859
Huber	-0.018	0.831
logistic	-0.018	0.835
Talwar	-0.017	0.787
Welsch	-0.017	0.791
OLS	-0.030	1.10
Theil's "incomplete"	-0.009	0.590
Theil's "complete"	-0.010	0.620

It is clear that POLDER has a negative offset, but the magnitude of the offset falls into three groups: OLS, robust WLS methods (first eight rows of table above) and robust non-parametric methods. Identical groupings of regression methods are found upon examining the slope values. Furthermore, omitting the small cluster of points with AERONET AOD>0.8, which were all measured on one day, namely 16 July 2012, the OLS slope becomes 0.7904 and offset is -0.014. This slope and offset are both very close to the slope and offset values from the various WLS fits. In the revised manuscript, we select to weight the fit residuals with Huber's function, for the following reason given by Bergstrom and Edlund (2013):

"while it still is robust, it does not completely disregard highly deviating points".

The table above shows that neither the offset nor the slope obtained with the Huber weights are outliers within the WLS group of robust regression methods. The tuning constant is assumed to be 1.345 following Holland and Welsch (1977).

At p3L29 of the revised manuscript, we now write:

Since individual AERONET and satellite AODs are not normally distributed, we use linear least-squares weighted by Huber's function to determine the slope and offset since this is a robust method that does not completely disregard highly deviating points (Bergström and Edlund, 2014). The slope and offset values determined using Huber's weighting function are encompassed by the values obtained with seven alternative weighting functions. Similarly, due to the non-normal distribution of the individual AOD data, Spearman's rank correlation coefficient (r_s) is chosen to study the site-specific AOD correlation based on individual AERONET-satellite coincidences.

References

Andrews, D. A., A robust method for multiple linear regression, Technometrics, 16(4), 523-531, 1974.

Bergström, P., and Edlund, O.: Robust registration of point sets using iteratively reweighted least squares, Comput. Optim. Appl., 58, 543–561, 2014.

Theil, H.: A rank-invariant method for linear and polynomial regression analysis: I, Proc. Kon. Ned. Akad. Wetensch., 53, 386-392, 1950.

Response to comments by reviewer 1

We thank the reviewer for sharing their expertise and improving the manuscript.

Major Comments:
As Andrew Sayer is an expert on aerosol retrievals from satellite-based remote sensing, I strongly recommend the authors fully take his suggestions.

We agree with this major comment and have taken most of the Dr. Sayer's suggestions.

Kahn et al., 2005 describes validation of a previous version of the algorithm and should be replaced with Kahn et al., 2010. The title is "Multiangle Imaging SpectroRadiometer global aerosol product assessment by comparison with the Aerosol Robotic Network". Particle mixtures have changed, but many of the notes the authors have made about MISR remain valid.

We have used the more recent reference suggested by the reviewer. We now write in Sect. 4:

The MISR low bias may be related to the need for darker spherical particles (Kahn et al., 2010) given that forest fire smoke plays a significant role throughout western Canada in the warm season (O'Neill et al., 2002). Spherical particles with lower single scattering albedo (SSA) may also be required to properly represent local anthropogenic pollution (Kahn et al., 2010) in the AOSR.

Although the paper is focused on AOD trends from satellite-remote sensing, I would recommend also including an analysis of the Fort McMurray AERONET site as well.

The AERONET data record is short (2005-2015) at Fort McMurray and includes a missing year (2006) and three currently incomplete years (2005, 2007, and 2015). The record effectively spans 2008 to 2014, which is too short for trend analysis, given the large interannual variability.

Page 8, Line 18-19: The higher SNR is probably irrelevant over land (especially bright surfaces).

Most of the retrieved AODs used in the temporal correlation with the Fort McMurray AERONET site, at least by MODIS, are over dark vegetation. However, SNR is valuable both for dark and bright scenes. To first order, the bright surface does not affect the number of detected aerosol-scattered photons, it essentially affects the number of photons reflected by the surface. So while a bright scene has less noisy radiances, the fractional contribution by aerosol scattering decreases relative to a dark scene and greater SNR is required to be able to detect a typical, small AOD (e.g. 0.1 at 550 nm) with comparable AOD precision relative to a dark scene. In spite of this point, we agree that the SNR of all instruments is probably sufficient and the higher SNR is likely irrelevant.
Thus, the relevant sentence in the manuscript becomes:

Stronger short-term correlation with AERONET AODs reflects the superior spatial resolution of the MODIS radiances (Table 1).

PM2.5 Assessment:
I strongly recommend that the authors remove the AOD-to-PM2.5 aspect of this paper. I don't think it adds much to the paper, as the authors have in-situ PM2.5 data for 10 sites anyways, and the correlation between AERONET AOD and satellite remote sensing retrieved AOD is much higher than the correlation between NAPS PM2.5 and satellite remote sensing retrieved AOD.

There are also a lot of caveats to converting between an integrated aerosol retrieval (AOD) and a surface aerosol retrieval (PM2.5), many of which I don't see discussed (please correct me if I missed it). Here are some of them:
1. For instance, MISR is viewing this area of the planet at roughly 10:15 AM local time. It is possible that the planetary boundary layer (PBL) is not always fully developed at this time, which would mean that a comparison between MISR AOD and surface based PM2.5 would not be possible.
2. Unmasked transported smoke that happens to be lofted above the PBL may not be seen by NAPS.
3. Variation in the PBL height from day to day and season to season will cause discrepancies between retrieved AOD and measured PM2.5 using a static ratio.
4. Large-scale differences in land-surface/water coverage may cause systematic discrepancies in PBL height at individual stations.
Although the results of the AOD-to-PM2.5 analysis show a positive trend in PM2.5 from space, I don't really see how useful this is, as the same thing can be shown from the 10 NAPS instruments with a much higher degree of confidence. Additionally, while I may trust the day-to-day changes in AOD retrieved from space, I would never put that kind of faith in converting AOD to PM2.5 on a daily basis. I recognize that the authors did not do this and are basically only using PM2.5 from AOD for yearly analysis, but some people may take this work and try to expand it in ways that probably shouldn't be done.

We agree with these comments. The AOD to $PM_{2.5}$ aspect can be avoided with the approach used in the revised manuscript. This involves correcting the POLDER/PARASOL and MODIS Deep Blue offsets (determined from the AERONET validation at Fort McMurray) and then calculating relative trends for AODs (from satellite) and for $PM_{2.5}$ (NAPS).

This is now described at p4L12:

For temporal trends, a simple linear regression is performed on relative anomalies derived from bias-corrected annual average and median AODs. The bias correction involves subtracting the AOD offset obtained through the validation with coincident Fort McMurray AERONET data.

General Comments:
Is it possible that the drop in 550 nm AOD (Figure 5) and NAPS PM 2.5 during 2015 is related to the fall in oil prices affecting activity in the region? If so, it may be worthwhile to note, as this would likely continue to the present day.

It is possible, but not likely, and this is too speculative in our opinion given that NAPS $PM_{2.5}$ data is not significantly different in 2013 and 2015 (see figure below illustrating oil prices over the past seven years). (http://www.nasdaq.com/markets/crude-oil.aspx?timeframe=7y).

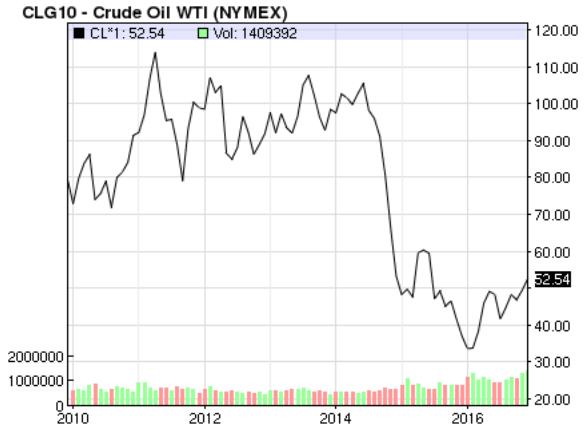


Figure 1: Figure 1 could be improved in a number of ways. In addition to what Andrew Sayer suggested, I recommend putting the locations of your AERONET sites and NAPS stations on the map (maybe as circles and stars). If you wanted to make the plots even more useful, you could color the circles and stars using the same color scale for AERONET, and a different scale for PM2.5.

The maps in Fig. 1 use all available satellite data, not just data that is coincident with $PM_{2.5}$ or AERONET observations. $PM_{2.5}$ is measured at night and in winter, when these satellite instruments do not measure. Similarly, the AERONET sites in the oil sands region measure all day, not just at the 1 or 2 local times per day of the satellite instruments and we have found diurnal variations in AOD of 30% at Fort McMurray based on AERONET data. Furthermore, AERONET has slightly more coverage during the cold season. To avoid these biases, in Fig. 1, we plot only the average AOD from satellite-coincident AERONET measurements. Both AERONET sunphotometers in the AOSR are collocated with NAPS sensors, so we chose AERONET over NAPS for Fig. 1. There is also the problem of a possible trend. The NAPS or AERONET data may cover a significantly shorter period (e.g. AERONET at Fort McKay started in 2013 whereas the POLDER map includes data from 1996). We leave Fort McKay out since the data record is too short for a reliable climatology of coincident AODs and has no temporal overlap with most of the sensors.

Figure 5: The authors should include the Fort McMurray AERONET site on this plot as well.

See response to earlier, related comment. No change is made to the manuscript.

Response to reviewer 2

We begin by thanking the reviewer for their very helpful comments.

However, I am concerns that this manuscript is insufficient to be useful due to lack of substantial materials and logical reasoning in current version.

We have added explanations to substantiate some of the results. For example, we now discuss a possible cause for why MISR does not capture the hotspot in climatological AOD as well as the other instruments. We have provided a reason why median AOD and $PM_{2.5}$ mass densities are preferable for the spatial correlation analysis in the revised manuscript (as opposed to mean values used in the original manuscript). We feel the discrepancy in long-term trends between the satellite sensors is not strong, but now suggest that the MODIS calibration degradation could account for the general negative trend in AOD from this sensor. Further details are provided below on each issue. This is simply a summary of our response.

First of all, I have read the comments from Andrew Sayer, who is an expert on aerosol retrievals from satellite-based remote sensing, especially in MODIS AOD retrievals. His comments are very useful to improve the understanding of the MODIS AOD retrievals and improve current studies.

Dr. Sayer's comments have helped to improve the revised manuscript. The reviewer can refer to our response to Dr. Sayer's comments to see the resulting changes to the manuscript.

My major concerns about this manuscript are the lack of in-depth analysis and lack the necessary explanations. For example, the finding of the ability to capture spatial variability with MISR is generally much worse than the other instruments over AOSR region is very interesting and useful to know the limitation of MISR measurements, however the possible reasons for this will be more important to see the spatial limitation of MISR.

The MISR spatial limitation, evident in Fig. 1, is probably due to its spatial sampling being tied to its temporal sampling. We found locations within the AOSR where MISR was measuring almost exclusively in October. Thus, the seasonal cycle in AOD is aliasing into the AOD spatial distribution.
The spatial correlation coefficient is based on 10 sites. Because of the small number of sites, the correlation is quite sensitive to a bias in AOD or $PM_{2.5}$ at any station. Wapasu has significantly higher mean $PM_{2.5}$ mass density for MISR coincidences than any other site (10.2 $\mu g/m^3$ while the next highest site average is 8.1 $\mu g/m^3$). MISR overpasses of Wapasu span only two years (2014-2015) and these years were affected by anomalously high forest fire activity in western Canada. The median reduces the sensitivity to these outliers as compared to the mean. In the revised manuscript, Table 3 now contains the correlation of the median of coincident $PM_{2.5}$ and satellite AOD data. This table is inserted below. The revised Table 3 shows the spatial correlation coefficient (R) of MISR AOD with $PM_{2.5}$ is not much worse than the spatial R of MODIS/Aqua DT and $PM_{2.5}$.

AOD product	R	N
POLDER/PARASOL 865 nm	0.64	8
AATSR 550 nm	0.73	9
MISR 558 nm	0.20	10
MODIS/Aqua DT 550 nm	0.23	10
MODIS/Aqua DB 550 nm	0.57	10

At p3L34, we now modify the description of the spatial correlation analysis as follows:

In order to assess the ability of the satellite data to capture the spatial variability in this region, the hourly in-situ surface-level $PM_{2.5}$ from the 10 NAPS (National Air Pollution Surveillance) stations (Table 2) are used. Demerjian (2000) provided a review of the NAPS network, but since 2011, this network has undergone a gradual shift in the continuous monitoring of $PM_{2.5}$ mass density from tapered element oscillating microbalances (TEOMs) to the SHARP (Synchronized Hybrid Ambient Real-time Particulate) monitoring system. The latter is a hybrid system, consisting of a nephelometer and a beta attenuation monitor (Hsu et al., 2016). The spatial correlation between median satellite AODs and NAPS $PM_{2.5}$ mass densities is determined using coincident data.

At p5L4, we now update the text with the following:

The AOD hotspot in the AOSR seen by POLDER is less obvious with MISR (Fig. 1). This is consistent with the relatively poorer ability of MISR to capture spatial variability based on spatial correlations of median AOD and $PM_{2.5}$ mass density over the ~10 available sites (Table 3). While the spatial correlation analysis relies on temporally coincident data, the less obvious AOD hotspot for MISR in Fig. 1 is also partly due to the spatio temporal sampling by this instrument. Some locations are only sampled during a short period of the year, and thus the seasonal cycle of AOD is aliased into the MISR spatial distribution.

In section 3.1, the authors have indicated that all of the satellite retrievals can capture the inter-annual variability of the annual mean AOD observed by AERONET, but the trends estimated based on the each satellite retrievals showed lots of differences, some of positive and some of negative. Thus, what are the main reasons to explain this discrepancy?

We agree that there is a discrepancy between the trends estimated by the different satellite AOD products, but it is not strong. The satellite data records all span approximately one decade. A period of a decade is rather short for determining a trend, considering the natural interannual variability in AOD and possible instrumental drifts (e.g. Levy et al., 2015). Focussing on the Muskeg River mining region where there appears to be a significant positive AOD trend according to MODIS/Aqua DB and POLDER/PARASOL, the linear trend is not different from zero for both AATSR and MISR (p6L28-29). Also, MODIS/Aqua DT has a slightly negative trend, but it is also not different from a null trend, so given that none of AOD products show a strong decreasing trend in this Muskeg River mining region, there is no strong discrepancy in the AOD trends. The insignificant negative AOD trend for MODIS/Aqua DT remains now that we have switched to 550 nm.

We now add at p8L1:

The calibration of the MODIS reflective solar bands is achieved by calibration with the solar diffuser. Some negative drift in AOD (Levy et al., 2015) is expected for MODIS Aqua similar to its Terra counterpart (see Sect. 2)

as the designs of the solar diffuser and its stability monitor are nearly identical in the two MODIS sensors (Wu et al., 2013).

The authors reported a major issue of satellite AOD retrievals over this region, which is the lack of successful retrieval samples, especially of the MODIS retrievals which has low confidence. It is good information. However, the reasons for the large part of retrievals has low confidence are not well explained.

The reasons for the low confidence of MODIS AODs were explained in the ACPD version of the manuscript (p5L24-26 for Deep Blue and p5L12-19 for Dark Target). An additional reason for MODIS Dark Target has been added to the revised manuscript: coastal areas (see comment by Dr. Sayer and response).

Furthermore, the comparison of coincident AODs observed by satellite-based and AERONET shows large bias (more than 20%) between them, but necessary explanations are not provided.

MISR is the only satellite-based aerosol sensors with a consistent bias of >20% in this region. Explanations were included in the ACPD version (p9L5-11), although one literature reference has been updated in these sentences.

I found that the correlation between monthly mean of the satellite retrieved AOD and AERONET AOD are analyzed, but I'd suggest to use the individual samples from AERONET to evaluate the satellite AOD retrievals and discuss the bias of each satellite product.

This is already done in Table 4. The second to fourth columns in Table 4, namely 'r_s', 'slope', and 'offset', are all based on individual coincidences. Although it can be inferred from the ACPD version of the manuscript that the quantities in these columns are based on a regression using individual coincidences (e.g. p1L12 and p9L3-4), we will be more explicit in Sect. 2. At p3L30, we now write

"Since individual AERONET and satellite AODs are not normally distributed, we use linear least-squares weighted by Huber's function to determine the slope and offset since this is a robust method that does not completely disregard highly deviating points (Bergström and Edlund, 2014). (…) Similarly, due the non-normal distribution of the individual AOD data, Spearman's rank correlation coefficient (r_s) is chosen to study the site-specific AOD correlation based on individual AERONET-satellite coincidences."

In the conclusion (p8L17), we now repeat that correlation was determined using individual AERONET observations:

"However, the MODIS dark target product is the best at capturing temporal variability in terms of the correlation with individual AERONET AODs at Fort McMurray…"

It is not clear to describe how to derive the PM2.5 mass density from satellite AODs. I noticed that the constant ratio of PM2.5 to AOD is used to convert the AOD trends from satellite instruments to PM2.5 trends. However, this is not accurate. The relationship between surface PM2.5 and AOD is not always linear. It is affected by multiple factors, such as the relative humidity since the AOD can be enhanced by aerosol swelling effects but the PM2.5 does not. Meanwhile, the correlation between AOD and surface level PM2.5 significantly depends on the

aerosol vertical distribution and aerosol particle size distribution. Thus, the uncertainties in those analysis and the influences on the results should be discussed.

The existing manuscript was not clear about the timescale when the word "constant" was used. What was meant is that the $PM_{2.5}$/AOD ratio is assumed to be constant from year to year (based on annually averaged ratios). This ratio can even change from year to year if there were an increasing trend in surface-level aerosol emissions. In the revised manuscript, we have devised a better way to compare trends: the POLDER/PARASOL and MODIS Deep Blue AOD offsets, determined from the AERONET validation at Fort McMurray, are corrected and then relative trends are used for $PM_{2.5}$ and satellite AOD. Thus, the $PM_{2.5}$/AOD ratio is not used in the revised manuscript. The Fort McMurray AERONET site is used for bias correction since it has temporal overlap with both sensors and has a longer record than the Fort McKay site. There is qualitative agreement on the magnitude of the offset at both sites for MODIS DB.

This is now described at p4L12:

For temporal trends, a simple linear regression is performed on relative anomalies derived from bias-corrected annual average and median AODs. The bias correction involves subtracting the AOD offset obtained through the validation with coincident Fort McMurray AERONET data.

P6, Line 28: Is this trend statistical significant?

Yes, the MODIS/Aqua DB and POLDER/PARASOL trends are both statistically significant. We will add "statistically" to the sentence as follows:

In fact, two satellite data products, namely POLDER/PARASOL and MODIS/Aqua DB, exhibit a statistically significant positive trend in this mining area.

Assessment of the aerosol optical depths measured by satellite-based passive remote sensors in the Alberta oil sands region

Christopher E. Sioris[1], Chris A. McLinden[1], Mark W. Shephard[1], Vitali E. Fioletov[1], and Ihab Abboud[1]

[1] {Environment and Climate Change Canada (ECCC), Toronto, ON, Canada}

Correspondence to: Christopher E. Sioris (christopher.sioris@canada.ca)

Abstract. Several satellite aerosol optical depth (AOD) products are assessed in terms of their data quality in the Alberta oil sands region. The instruments consist of MODIS (Moderate resolution Imaging Spectroradiometer), POLDER (Polarization and Directionality of Earth Reflectances), MISR (Multi-angle Imaging SpectroRadiometer), and AATSR (Advanced Along-Track Scanning Radiometer). The AOD data products are examined in terms of multiplicative and additive biases determined using local AERONET (AEROCAN) stations. Correlation with ground-based data is used to assess whether the satellite-based AODs capture day-to-day, month-to-month, and spatial variability. The ability of the satellite AOD products to capture interannual variability is assessed at Albian Mine and Shell Muskeg River, two neighbouring sites in the northern mining region where a statistically significant positive trend (2002-2015) in $PM_{2.5}$ mass density exists. An increasing trend of similar amplitude (~5%/year) is observed in this northern mining region using some of the satellite AOD products.

1 Introduction

Fine-mode aerosols can be harmful to the respiratory system in large doses and are thus a critically important constituent with regard to air quality. For this reason, particulate matter with median aerodynamic diameter less than 2.5 μm ($PM_{2.5}$) is one of the atmospheric observables used to calculate the Air Quality Health Index (AQHI) in Canada (Stieb et al., 2008). Similar indices are used in other countries (Kelly et al., 2012). Tropospheric aerosols are also a major source of uncertainty in estimating the radiative forcing of climate (Myhre et al., 2013). Many satellite-based instruments can provide information about atmospheric aerosols in the form of aerosol optical depth (AOD), a measure of the vertically integrated extinction of the solar beam by aerosols. Measurements of AOD tend to be proportional to particulate matter mass density measured at the surface when the boundary layer aerosol concentrations are elevated (e.g. Tian and Chen, 2010).

The Alberta oil sands region (AOSR) has been under rapid industrial development during the past decade (Foote, 2012). Satellite measurements already indicate a significant increasing trend in nitrogen dioxide between 2005 and 2014 (McLinden et al., 2012; McLinden et al., 2016). Additionally, the AOSR is being deforested as part of expanding surface mining operations. This inevitably increases levels of dust, which partly arises from

transportation by trucks. Dust is one of many aerosol types of relevance in the AOSR. Other main aerosol types include organic aerosols, both natural and anthropogenic (Liggio et al., 2016), as well as ammonium sulfate.

Passive remote sensing of aerosol over land is challenging because, for a cloud-free scene, most of the nadir radiance is coming from direct reflection off the surface at visible wavelengths, not from aerosol scattering. This is particularly true for the AOSR, which consists of an irregularly-shaped industrial area to the south comprised of non-vegetated (cleared) mining locations and a second area to the north where mostly surface mining is occurring, as both areas have high surface albedo in the visible. Within the AOSR, the land type changes on spatial scales smaller than the typical 10 × 10 km AOD footprint of a satellite-based instrument. Considering the area surrounding the AOSR, specifically the rectangular area between 55.0 and 58.5°N and 114.0 to 108.5°W, the land is covered by evergreen needleleaf forest (70%) and some deciduous broadleaf forest (23%), which is typical of the boreal forest in the northern portions of the Alberta and Saskatchewan.

2 Method

In order to study the spatiotemporal distribution of AOD in the AOSR, data from several satellite-based instruments are used. Satellite-based aerosol sensors are chosen based on a number of factors. One of the goals of the study is to examine long-term AOD trends, so preference is given to instruments with longer data records. Instruments that view a scene with multiple viewing angles were selected as the multi-angle capability is useful for disentangling the contributions to the scene reflectance by the surface and by the overlying aerosols (e.g. Bevan et al., 2012). Such instruments include Multi-angle Imaging SpectroRadiometer (MISR) (Diner et al., 1989), the Polarization and Directionality of Earth Reflectances (POLDER) series (Deschamps et al, 1994) including POLDER/PARASOL (Polarization & Anisotropy of Reflectance for Atmospheric Sciences coupled with Observations from a Lidar), and the Along-Track Scanning Radiometer (ATSR) series (see Table 1 for the spatial resolution, temporal coverage and wavelength at which AOD is reported for each of the satellites). In addition, MODIS (Moderate resolution Imaging Spectroradiometer) is chosen partly because it has a long wavelength channel (2.1 μm) that allows the surface reflectance to be accurately determined over vegetation without contamination from fine-mode aerosols (e.g. particles with radii of <0.2 μm) by virtue of the correlation between visible and 2.1 μm surface reflectance for vegetation (e.g. Kaufman et al., 2002; Li et al., 2005). MODIS/Aqua collection 6 data are used (see Appendix for providers and version numbers of other satellite data products). For MODIS, there are two AOD retrieval algorithms yielding the Dark Target (DT) (Levy et al., 2013) and the Deep Blue (DB) (Hsu et al., 2013) products. Specifically, the Corrected_Optical_Depth_Land (~~470~~550 nm) and the Deep_Blue_Aerosol_Optical_Depth_550_Land datasets were used and confidence for both datasets was extracted from the Quality_Assurance_Land dataset. The Dark Target algorithm exploits the fact that, for dark surfaces, aerosols tend to brighten the scene. For highly reflective surfaces such as snow in the visible spectral region, AOD cannot be retrieved using either the DT or DB approach. The MODIS Aqua DT product is also processed at 3 km spatial resolution in addition to the standard 10 km resolution available for both MODIS products (Levy et al., 2013). Each MODIS AOD measurement is assigned a confidence value. Confidence values of 1 and 0 indicate marginal and no confidence, respectively, while values of 2 and 3 represent good and ideal confidence (Levy et al., 2013). For MODIS/Aqua collection 6, data with

confidence≥1 are retained for validation. The theoretical basis of the MISR aerosol retrieval algorithm is given by Diner et al. (2008). The aerosol retrieval for AATSR is described by Bevan et al. (2012) and references therein. Deuzé et al. (2001) detail the approach used to retrieve aerosol information from POLDER observations over land.

MODIS Terra is not considered since it is highly similar to MODIS Aqua but, for collection 6, the former is less reliable for trend studies in spite of improvements relative to collection 5 (Levy et al., 2015). The MODIS-based Multiangle Implementation of Atmospheric Correction (MAIAC) (Lyapustin et al., 2011) product is not currently available in the AOSR (van Donkelaar et al., 2016). VIIRS (Visible Infrared Imaging Radiometer Suite) (Hillger et al., 2013) is not considered in this study because of its shorter data record relative to the MODIS sensors. Active remote sensing instruments are not considered because of the long revisit time and poor spatial coverage of the relatively small AOSR. The focus in this paper is primarily on the different aerosol sensors, rather than the different retrieval algorithms applied to the same satellite data (e.g. Popp et al., 2016), with the exception of the widely used Deep Blue and Dark Target algorithms for MODIS.

For validation of satellite-based AOD data, AERONET (Holben et al., 1998) is the ideal choice since the same quantity is measured by this ground-based network of direct-sun multiband photometers and the ~3 minute typical sampling interval generally ensures a good temporal coincidence during clear sky conditions. Quality-controlled AERONET data (Level 2, version 2) are used (http://aeronet.gsfc.nasa.gov). ~~CIMEL~~Cimel (French manufacturer) CE318 sensors used by AERONET measure at several wavelength, some of them (e.g. 500 and 870 nm) are close to the wavelengths at which the selected satellite instruments report AOD (e.g. 470, ~550, and 865 nm). There are two AERONET sites in the oil sands region: Fort McMurray (56.752°N, 111.476°W) and Fort McKay (57.184°N, 111.64°W). Measurements at Fort McMurray started in 2005. The Fort McKay site has only been in operation since August 2013 meaning that there is no temporal overlap with Advanced ATSR (AATSR) and only seven coincidences with POLDER/PARASOL using coincidence criteria of ±12 minutes and 10 km. The spatial coincidence criterion corresponds to the smallest AOD footprints of the selected data sets (Table 1). A larger spatial coincidence criterion is not considered since, as shown below, strong spatial gradients in AOD exist in this aerosol source region. Furthermore, as mentioned in Sect. 1, the surface type also changes on such spatial scales. The temporal coincidence criterion was set to limit the number of independent AERONET measurements used in the statistical analysis. There can be multiple AERONET observations that are temporally coincident with a satellite observation and there can be up to four spatial coincident satellite AODs during a satellite overpass of an AERONET site. All of these coincidences are treated as independent data points in the validation and correlation analyses. In order to properly validate satellite AOD bias, AERONET 500 nm AODs are interpolated to the satellite AOD wavelengths (see Table 1) using the coincident AERONET Ångström exponent derived from 440 and 675 nm measurements, except for POLDER/PARASOL, for which no scaling of the AERONET AOD was applied. Since individual AERONET and satellite AODs are not normally distributed, we use linear least-squares weighted by Huber's function to determine the slope and offset since this is a robust regression method that does not completely disregard highly deviating points (Bergström and Edlund, 2014). The slope and offset values determined using Huber's weighting function are encompassed by the values obtained with seven alternative weighting

functions. Similarly, due the non-normal distribution of the individual AOD data, Spearman's rank correlation coefficient (r_s) is chosen to study the site-specific AOD correlation based on individual AERONET-satellite coincidences.

The ability of each satellite-based sensor to capture the AOD seasonality in snow-free months is determined at Fort McMurray using ~~the~~Pearson's correlation of monthly averaged AODs (using all overlapping years) with AERONET. A minimum of 20 coincident data points per calendar month must be available for that month to be included in the correlation.

In order to assess the ability of the satellite data to capture the spatial variability in this region, ~~spatial correlation is determined for~~the hourly in-situ surface-level $PM_{2.5}$ from the 10 NAPS (National Air Pollution Surveillance) stations (Table 2) ~~and satellite AODs averaged over all coincidences within their temporal overlap period. NAPS stations continuously monitor~~are used. Demerjian (2000) provided a review of the NAPS network, but since 2011, this network has undergone a gradual shift in the continuous monitoring of $PM_{2.5}$ mass density ~~using~~from tapered element oscillating microbalances (TEOMs~~). The NAPS network is reviewed by Demerjian (2000), although recently there has been a gradual shift in technology since 2011~~) to ~~a~~the SHARP (Synchronized Hybrid Ambient Real-time Particulate) monitoring system~~, which~~. The latter is a hybrid system, consisting of a nephelometer and a beta attenuation monitor (Hsu et al., 2016). The spatial correlation between median satellite AODs and NAPS $PM_{2.5}$ mass densities is determined using coincident data. The use of medians rather than means reduces the sensitivity to outliers from forest fires. The same 10 km spatial coincidence criterion is used but temporal coincidence limit is extended to ±1 hour to match the temporal resolution of the selected NAPS datasets.

Similar to the spatial and seasonal variability, the ability of the satellite instruments to capture interannual variability can be assessed by correlating yearly satellite-based AODs averaged over all coincidences with NAPS $PM_{2.5}$ measurements over the overlap period. 20 coincidences in a calendar year are required for the year to be included in the correlation calculation. As an example, for MISR, 14 sufficiently sampled years (2002-2015) are used in the correlation with NAPS data at Millennium mine.

For temporal trends ~~in AOD, a~~, an ordinary least-squares simple linear regression is performed on relative anomalies derived from bias-corrected annual average and median AODs. The bias correction involves subtracting the AOD offset obtained through the validation with coincident Fort McMurray AERONET data. The mean of the yearly averages ~~and~~or medians is used to compute the relative anomalies. Similarly, for $PM_{2.5}$, the annual average of daily average values are used since the $PM_{2.5}$ auto-correlation timescale is on the order of 6.5 hours, based on analysis of Albian mine $PM_{2.5}$ data from 2002. The extra step of daily averaging prior to annual averaging yields more conservative annual standard error (s. e.) estimates. Partial years at the start and the end of a data record are removed. Trend periods are given below for each sensor. The area over which the satellite-based AOD trend maps are calculated is 0.1°×0.1° by default. This default setting is used to determine the AOD trend for both MODIS/Aqua 10 km products (2003-2015). The trend domain considered in this work spans from 56-58°N and 111-112°W. For sensors with poorer spatial coverage (MISR, AATSR, POLDER/PARASOL), the spatial binning is expanded in latitudinal and longitudinal increments of 0.1° until there are ≥20 observations in each calendar year

within at least one grid cell in the domain. The trend maps are ultimately generated at 0.3°×0.3° for AATSR (2003-2011) and MISR (2000-2015) whereas a 0.4°×0.4° area is required for POLDER/PARASOL (2005-2013). Outlying individual data points (>4 standard deviations above the climatological average in the domain) are recursively filtered mainly to reduce the influence of forest fires on trends. The same filtering is applied to the $PM_{2.5}$ datasets. Interannual consistency in the month-to-month sampling is checked for any location with a positive satellite AOD trend significant at the 95% confidence interval by calculating the average day-of-the-year for each calendar year. Such temporal sampling anomalies occur for MISR AOD data at some locations if a 0.1°×0.1° grid were used, for example. The Albian mine (2001-2008) and Shell Muskeg River (2009-2015) forest-fire-filtered $PM_{2.5}$ datasets were merged for trend analysis since the sensor was relocated from the former to the latter site in January 2009 and these sites are separated by <5 km.

3 Results

First, the general spatial distribution of AOD is illustrated for some of the aforementioned data sets. In Fig. 1, the climatological average POLDER AOD on a 0.1° × 0.1° grid is shown. This is the default grid used for climatological maps of all satellite AOD datasets. The POLDER sample size per grid cell is 90 to 170 in the AOSR over the discontinuous period from 1996 to 2013 (see Table 1). There is a clear hotspot in 865 nm AOD in the AOSR region, roughly double the surrounding background values. Note that for POLDER and MISR, there are expected voids in their spatial coverage (Fig. 1) due to the spatial sampling of these instruments, whereas MODIS and AATSR footprints can be centered on any geolocation within the AOSR.

The AOD hotspot in the AOSR seen by POLDER is less obvious with MISR (Fig. 1). ~~The~~This is consistent with the relatively poorer ability of MISR to capture spatial variability ~~with MISR is generally much worse than the other instruments~~ based on spatial correlations of ~~average satellite-based~~median AOD ~~versus average NAPS~~and $PM_{2.5}$ mass density over the ~10 available sites (Table 3). While the spatial correlation analysis relies on temporally coincident data, the less obvious AOD hotspot for MISR in Fig. 1 is also partly due to the spatiotemporal sampling by this instrument. Some locations are only sampled during a short period of the year, and thus the seasonal cycle of AOD is aliased into the MISR spatial distribution. Table 4 provides the number of coincidences for each satellite with the both Fort McMurray and Fort McKay AERONET observations to provide a ~~relative~~ sense of how the coincident sample sizes vary as a function of the satellite AOD data product.

The climatological AOD maps for the MODIS/Aqua collection 6 DT and DB products (2002-2014) are also shown in Fig. 1 however there is a major issue with the confidence as shown in Fig. 2. Near the Syncrude facility at Mildred Lake (57.05°N, 111.6°W), the confidence approaches 0 in both MODIS products in the two adjacent 0.1°×0.1° cells (Fig. 2). In the western cell, the inadequate confidence in MODIS Aqua collection 6 DT data is due to failure of the AOD retrieval algorithm due to the 2.1 μm reflectance exceeding the allowed upper limit of 0.35. This is a fundamental ~~weakness~~limitation of the Dark Target retrieval strategy (see sect. 2). In the adjacent eastern cell, the low confidence stems from the low number of 0.5×0.5 km^2 pixels (see Table 1) used in the AOD retrieval. The number of pixels used in the AOD retrieval is reduced by the inland water mask (Carroll et al., 2016), the high

2.1 μm reflectance (>0.35), but also by cloud masking and an independent test for optically thicker cirrus, diagnosed using the 1.38 μm channel (Levy et al., 2013; Hubanks, 2015). The high reflectance in the near-infrared affecting the western cell and possibly the eastern cell is typical of desert or sandy loam. The higher spatial resolution of the MODIS-Aqua 3 km DT data clarifies the importance of this issue: key areas in the AOSR are simply not monitored with confidence by the current MODIS/Aqua DT product. For example, there are 0.01° × 0.01° areas with no AOD measurements of the highest confidence in 12 years, whereas surrounding, equal areas have tens of observations. The lack of confidence is not unique to the AOSR. Low confidence is also observed in urban areas within the province (e.g. Calgary, not shown). The low confidence in the MODIS DB product is due to the spatial heterogeneity of the surface between vegetated and non-vegetated area, which leads to pixels falsely identified as cloudy (N. Christina Hsu, NASA, priv. communication). Li et al. (2009) identified the need for improved AOD measurements using the DB algorithm over transitional land covers.

A similar issue exists for AATSR (Fig. 3) and ATSR-2 (not shown), which both have an exceedingly small number of successful retrievals in a 0.1° × 0.1° area containing the Mildred Lake Syncrude facility (e.g. N<10) during their respective missions (Table 1). Similarly to MODIS, this is probably caused by falsely identifying bright patches in otherwise vegetated scenes as clouds (P. North, Swansea University, priv. communication). Cloud fraction for successful AOD retrievals tends to be as high as 0.18 within the oil sands region, including the northern mining region, yet drops to 0.02 in the surrounding region (Fig. 3). Note that cloudy 1×1 km^2 pixels are not used during the AATSR AOD retrieval. The spatial correlation coefficient between sample size and cloud fraction as illustrated in Fig. 3 is -0.73, indicating that the spatial variation in AATSR sample size is mostly related to cloud flagging. Neither POLDER nor MISR show a sampling void in the AOSR. Table 1 shows that these two sensor types have coarser AOD spatial resolution by a factor of 3-4 than MODIS, ATSR-2, and AATSR. Note that some of the $PM_{2.5}$ sites are located in the periphery of the industrial and mining areas and thus spatial coincidences exist for MODIS and AATSR in spite of the aforementioned issues, given the 10 km coincidence criterion.

In terms of the validation using AERONET data (Table 4), MISR has a ~~large~~ multiplicative bias (i.e. ~~small~~ slope significantly less than unity), which is ~~consistent between~~similar at both sites in the AOSR. ~~Excluding Fort McMurray coincidences for which the AERONET AODs interpolated to 558 nm are >0.4, the~~The slope ~~improves to 0.74 and~~ is of a similar value to the slope found in previous studies for inland (Liu et al., 2004), dusty (Kahn et al., 2005), and urban environments (Jiang et al., 2007). MODIS DB tends to yield more data than the DT product, but the correlation is lower with AERONET on individual coincidences and in terms of the seasonal variation. At ~~both~~the Fort McMurray AERONET sites, the MODIS products behave oppositely in terms of multiplicative and additive biases (discussed in Sect. 4). AATSR and POLDER/PARASOL ~~show no major deficiencies, with the latter exhibiting the closest slope value to unity of all of the satellite sensors at Fort McMurray.~~have no major validation shortcomings.

3.1 Trends

Before considering trends in the AOSR, it is useful to look at whether the different satellite data products capture the AOD interannual variability at Fort McMurray, where a sufficiently long record (2005-2015) of 500 nm AOD exists. All of the products capture the interannual variability of the annual mean AOD observed by AERONET at Fort McMurray (Table 5). Correlation coefficients for forest-fire-filtered annual means tend to be only slightly lower.

In general, very few of the 200 grid cells in the trend domain (56-58°N, 111-112°W) indicate a statistically significant (2 s. e.) positive trend that is consistent from one satellite to the next. In fact, there are no points in the domain for which MODIS/Aqua DT (2003-2013), AATSR, or ATSR-2 (1996-2002, 0.3°×0.3°) show a significant positive trend in AOD. Similarly, POLDER/PARASOL only shows a significant positive trend in three adjacent grid points at 57.3°N between 111.3 and 111.5°W (see Fig. 4) and MISR also finds a significant positive trend at only two locations in the domain. Finally, MODIS/Aqua DB has two points with the largest and most significant positive AOD trend in the region of the Muskeg River mine at 57.25°N, 111.25°W (Fig. 4). In fact, two satellite data products, namely POLDER/PARASOL and MODIS/Aqua DB, exhibit a statistically significant positive trend in this mining area. Although not statistically different from zero, the AOD trend in both AATSR and MISR data is positive in the area of the positive POLDER/PARASOL trend (Fig. 4), whereas MODIS DT tends to show an insignificant negative trend.

Changes to the surface may be at the root of the increasing AOD trend in this area, either since clearing of vegetation could lead to higher concentrations of dust, or by biasing the AOD retrieval. Trends in surface albedo were determined from the combined MODIS Terra/Aqua MCD43C3 albedo data product at four wavelengths relevant to the MODIS or POLDER AOD retrievals: 470, 645, 860, and 2130 nm (see Appendix A). For all four wavelengths, neither the largest nor the most significant trends in surface reflectivity occur at 57.25°N, 111.25°W (not shown), where the largest and most significant MODIS DB AOD trend occurs and also within the larger area of the spatially coherent POLDER/PARASOL AOD trend.

~~In order to quantitatively compare trends in AOD and $PM_{2.5}$, the ratio of the average AOD to average $PM_{2.5}$ mass density over all coincidences between each satellite instrument and a given NAPS site is used to convert the AOD trends from the satellite instruments to $PM_{2.5}$ trends. This implicitly assumes that the ratio of $PM_{2.5}$ to AOD is constant over time. This ratio is determined for the merged Albian mine / Shell Muskeg River dataset.~~ Since aerosol optical depth histograms indicate a skewed distribution, it is ~~also~~ useful to verify trends using annual medians. ~~For that purpose, the ratio of median AOD to median $PM_{2.5}$ is used instead. This approach is particularly important for POLDER/PARASOL because of the very low 865 nm AODs (Fig. 1) and the negative offset (Table 4) that do not allow a relative trend to be meaningful.~~

A significant positive trend of ~~0.24±0.06~~(4.1±1.1)%/year (±1 standard error) (Figs. 5-6) and ~~0.24±0.07 μg/m³/~~(5.7±1.6)%/year is detected in the Albian mine/ Shell Muskeg River merged annual average and median $PM_{2.5}$ mass densities (2002-2015), respectively. Limiting the merged $PM_{2.5}$ dataset to the warm season (April-October) to mimic the temporal coverage of the satellite data (Table 4), the ~~trend (0.25±0.07 μg/m³/~~relative trend

using annual averages is (4.6±1.2)%/year~~) does~~, which is not change significantly from the trend using year-round data (Fig. 6). A consistent trend of ~~0.21±0.09 μg/m^3/~~(4.8±2.1)%/year is found in annually-averaged $PM_{2.5}$ at Albian mine (2002-2008) alone, and the trend there during the warm season is (4.3±1.1)%/year, also statistically significant and not different ~~(0.24±0.06 μg/m^3/~~from the year~~),~~-round trend. Furthermore, there is no indication of a discontinuity between 2008 and 2009 when the monitoring site was relocated. The relative trend in $PM_{2.5}$ at the surface is in quantitative agreement with the ~~$PM_{2.5}$~~ relative trends derived from MODIS/Aqua Deep Blue and POLDER/PARASOL annually averaged AOD data over similar, yet shorter periods. For both MODIS/Aqua Deep Blue and POLDER/PARASOL, trends using annual medians agree with trends determined using annual averages within their respective standard errors (1 s. e.). The low bias of POLDER/PARASOL AOD near these two Shell mines is expected from the validation with AERONET at Fort McMurray (Table 4) and previous work on larger spatial scales (Deuzé et al., 2001).

Contrary to the localized, significant AOD trend in satellite data records in the eastern portion of the Muskeg River region, a statistically significant trend is found at two other ground-based stations within the AOSR for the period 2002-2014, namely Syncrude UE1 and Millennium mine (Fig. 6). The largest trend occurs at Millennium mine, the closest NAPS station to the southeast of the Shell Muskeg River region (see Table 2 and Fig. 4 for location). The trend is insignificant using either annual means or median $PM_{2.5}$ data at CNRL Horizon and Anzac where data records are shorter, while the trend at Wapasu (2013-2015) was not evaluated. The $PM_{2.5}$ trends at the remaining sites in the AOSR, namely two sites at Fort McMurray and one at Fort McKay are discussed below. Note that POLDER/PARASOL does not measure at Syncrude UE1 (see Table 3) and there is insufficient sampling at Millennium Mine over an area of 0.4°×0.4° in each of the years (2005-2013) for trend analysis. For POLDER/PARASOL, the trend, while mostly insignificant in the AOSR, is always positive. For AATSR, the AOSR has regions of statistically insignificant negative and positive trends. For MISR, the trend is positive in 56% of the trend domain and even more so (83%) in the northern half of the domain (57-58°N). For MODIS DB and DT, some of the AOSR is not sufficiently sampled with high confidence (see Sect. 2), but where confidence is ≥1, the trend tends to be negative in 69% and 77% of this area, respectively. The calibration of the MODIS reflective solar bands is achieved by calibration with the solar diffuser. Some negative drift in AOD (Levy et al., 2015) is expected for MODIS Aqua similar to its Terra counterpart (see Sect. 2) as the designs of the solar diffuser and its stability monitor are nearly identical in the two MODIS sensors (Wu et al., 2013). Li et al. (2016) find a small positive trend in AOD over Athabasca (56-58°N, 110-113°W) using MODIS/Aqua DB data (2004-2015), insignificant at the 2 s. e. level. Bari and Kindzierski (2016) found no indications of a positive trend in $PM_{2.5}$ at Fort McKay and the Fort McMurray Athabasca Valley site, using a longer period (1998-2014), although, as shown in Fig. 2 of Bari and Kindzierski (2016) for Fort McKay, there is an abrupt decrease in $PM_{2.5}$ mass densities that occurs between 2001 and 2002 that has a profound effect on the trend and its uncertainty. This discontinuity is observed at all sites in the AOSR that extend back to 2001. An earlier study by the same authors (2015) also indicated no trend between 1998-2012 at the same sites and at the Fort McMurray Patricia MacInnes site as well. ~~Li et al. (2016) find a small positive trend in AOD over Athabasca (56-58°N, 110-113°W) using MODIS/Aqua DB data (2004-2015), insignificant at the 2 s. e. level.~~

4 Discussion and conclusions

In this section, the advantages and limitations of the various data products are summarized. As shown in Table 4, all of the satellite sensors capture the temporal variability in AOD over Fort McMurray, based on correlations with AERONET, in spite of the low AODs there (e.g. Fig. 1). This temporal variability is largely driven by day-to-day variability as forest fires lead to episodes with large AODs (>3) in summer months that strongly influence the calculated correlation.

The two MODIS AOD data products (Deep Blue and Dark Target) have low confidence in the AOSR partly due to issues relating to elevated surface reflectivity in the vicinity of the Mildred Lake Syncrude facility. However, the MODIS dark target product is the best at capturing temporal variability in terms of the ~~correlations~~correlation with individual AERONET ~~AOD~~AODs at Fort McMurray and in terms of capturing the month-to-month variability. ~~This is likely due to MODIS's combination of~~ Stronger short-term correlation with AERONET AODs reflects the superior spatial resolution of the MODIS radiances (Table 1~~)~~), which is useful for resolving and ~~higher signal-to-noise ratio (SNR): its radiances have SNR > 1000 (Xiong et al., 2003) whereas the other instruments have SNR~~filtering small clouds and localized areas of ~~1000 or less (Deschamps et al., 1994; European Space Agency, 2007; Diner et al., 1989). MODIS DT clearly has a slope slightly greater than unity over the AOSR, in contrast to MODIS DB (Table 4).~~ high surface reflectively. Focussing on Fort McMurray, where there is a longer AERONET data record than at Fort McKay, ~~the~~ MODIS DT has a slope ~~changes insignificantly when coincident AERONET AOD is limited~~greater than unity, in contrast to ~~<0.7~~MODIS DB (Table 4). The same pattern of ~~consistently~~ high and low slope values for the MODIS Aqua DT and DB (collection 6) products, respectively, was found over two sites in Pakistan, namely Lahore and Karachi, by Bilal et al. (2016) and during non-fire summertime periods over semi-arid Nevada and California as shown in Table 4 of the work of Loría-Salazar et al. (2016). A high slope may be related to the use of the 2.1 µm channel to determine the reflectivity in the visible over non-vegetated surfaces as suggested by Bilal et al. (2016). High-biased AODs result because the surface reflectance in the visible assumed by the retrieval algorithm is less than the actual value as the relationship between the visible and 2.1 µm was developed for vegetated land for which a stronger spectral variation exists than for barren land. Li et al. (2005) have shown that the spectral reflectance relationship is much different even for dry vegetation than green vegetation. Note that high day-to-day variability can be captured in spite of biases in assumed surface reflectance since the latter changes slowly with time over the warm season, when successful measurements occur more frequently. A MODIS algorithm designed to function over inhomogeneous surfaces such as the AOSR region, and which would also likely be applicable to urban areas, is being investigated to exploit the many benefits of MODIS radiance data. One such benefit is the twice-daily revisit over the AOSR that the current multi-angle sensors, namely MISR and SLSTR (Sea and Land Surface Temperature Radiometer) (Coppo et al., 2010), cannot offer. SLSTR, onboard the recently launched Sentinel-3a satellite, is the next generation in the ATSR series.

MISR clearly captures the short-term and month-to-month AOD variability at Fort McMurray based on correlations at the individual coincidence level and the monthly time scale (Table 4), but struggles to capture the local spatial variability including the AOD hotspot in the AOSR as discussed in Sect. 3. The MISR low bias may be related to the

need for darker spherical particles (Kahn et al., ~~2005~~2010) given that forest fire smoke plays a significant role throughout ~~the~~ western Canada in the warm season (O'Neill et al., 2002). Spherical particles with lower single scattering albedo (SSA) may also be required to properly represent local anthropogenic pollution (Kahn et al., ~~2005~~2010) in the AOSR. The 3×3 superpixel averaging that is used when the MISR retrieval fails for the central superpixel could also contribute to a low bias (Jiang et al., 2007), particularly at Fort McKay as background AODs to the west could be lowering the average.

AATSR has a major spatial sampling issue in the heart of the AOSR, but also captures month-to-month variability from late spring to early autumn (Table 4) as well as short-term (Table 4) and spatial variability (Table 3). Based on a previous analysis (Che et al., 2016), the AATSR AOD underestimation of the Swansea University product (also used here) is larger over barren surfaces or sparse vegetation. Such land cover types are present in the AOSR. ~~The slight bias (Table 4) is not strongly AOD-dependent as removing coincidences with AERONET 500-nm AOD of ≥0.35 does not significantly change the slope of the regression equation (Table 4).~~

POLDER has a known negative offset in AOD (Deuzé et al., 2001), confirmed using coincident Fort McMurray AERONET AOD data. ~~However, POLDER/PARASOL is~~ For the temporal trend calculation, the approach of using relative anomalies based on bias-corrected AODs is particularly important for POLDER/PARASOL because the very low 865 nm AODs (Fig. 1) and the negative offset (Table 4) do not allow a relative trend to be meaningful without bias correction. Nevertheless, POLDER/PARASOL is the among the most accurate satellite-based aerosol sensor at Fort McMurray during periods of the higher AODs (e.g. ≥0.~~31~~4, Table 4), when its negative offset becomes rather trivial. Overall, the POLDER AOD product is without a major weakness relative to the other instruments, although it is provided at a relatively coarse spatial resolution (Table 1) and the fixed spatial sampling pattern of this sensor inhibits the application of spatial oversampling techniques. The use of polarized radiances reduces the sensitivity of the retrieved AOD to surface reflectance (e.g. Deuzé et al., 2001). The trend in POLDER/PARASOL AOD at the Shell mines (Albian and Shell Muskeg River) is probably not driven by a trend in surface reflectance since agreement with AERONET tends to be independent of surface type (e.g. Chen et al., 2015). A future sensor of POLDER heritage, namely the Multi-viewing, Multi-channel, Multi-polarisation Imager (3MI), offers higher spatial resolution, the availability of longer wavelength channels, and the potential for accurate monitoring of the local aerosol loading in the decade to come.

While AODs in the AOSR are relatively small according to POLDER/PARASOL (Fig. 1), the significantly positive trend in AOD from this satellite sensor and the similar trend in observed surface-level $PM_{2.5}$ in the region of the Muskeg River mine points to the need to continue monitoring of this region with a combination of surface and satellite-based aerosol observations.

Acknowledgements. Helpful discussions with Shailesh Kharol (ECCC) on the size range of dust particles are gratefully acknowledged. The European Space Agency Climate Change Initiative program is acknowledged. Peter North (Swansea University) is thanked for comments on an earlier version of the manuscript.

5 References

Bari, M., and Kindzierski, W. B.: Fifteen-year trends in criteria air pollutants in oil sands communities of Alberta, Canada, Environment International, 74, 200–208, 2015.

Bari, M. A., and Kindzierski, W. B.: Evaluation of air quality indicators in Alberta, Canada – An international perspective, Environment International, 92-93, 119-129, 2016.

Bergström, P., and Edlund, O.: Robust registration of point sets using iteratively reweighted least squares, Comput. Optim. Appl., 58, 543–561, 2014.

Bevan, S. L., North, P. R. J., Los, S. O., Grey, W. M. F.: A global dataset of atmospheric aerosol optical depth and surface reflectance from AATSR, Remote Sens. Environ., 116, 199–210, 2012.

Bilal, M., Nichol, J. E., and Nazeer, M.: Validation of Aqua-MODIS C051 and C006 operational aerosol products using AERONET measurements over Pakistan, IEEE J. Selected Topics Appl. Earth Observations Remote Sens., 9, 2074-2080, 2016.

Bréon, F. M.: Parasol Level-2 product data format and user manual, Ed. 1 – Rev. 6, 2011.

~~Che, Y., Xue, Y., Mei, L., Guang, J., She, L., Guo, J., Hu, Y., Xu, H., He, X., Di, A., and Fan, C.: Technical note: Intercomparison of three AATSR Level 2 (L2) AOD products over China, Atmos. Chem. Phys., 16, 9655–9674, 2016.~~

Carroll, M. L., DiMiceli, C. M., Townshend, J. R. G., Sohlberg, R. A., Elders, A. I., Devadiga, S., Sayer, A. M., and R. C. Levy, R. C.: Development of an operational land water mask for MODIS Collection 6, and influence on downstream data products, Int. J. Digital Earth, doi: 10.1080/17538947.2016.1232756, 2016.

Chen, H., Cheng, T., Gu, X., Li, Z., and Wu, Y.: Evaluation of polarized remote sensing of aerosol optical thickness retrieval over China, Remote Sens., 7, 13711-13728, doi:10.3390/rs71013711, 2015.

Coppo, P., Ricciarelli, B., Brandani, F., Delderfield, J., Ferlet, M., Mutlow, C., Munro, G., Nightingale, T., Smith, D., Bianchi, S., Nicol, P., Kirschstein, S., Hennig, T., Engel, W., Frerick, J., and J. Nieke: SLSTR: a high accuracy dual scan temperature radiometer for sea and land surface monitoring from space, J. Modern Opt., 57(18), 1815-1830, doi:10.1080/09500340.2010.503010.

Demerjian, K. L.: A review of national monitoring networks in North America, Atmos. Environ., 34, 1861-1884, 2000.

Deschamps, P.-Y., Bréon, F.-M., Leroy, M., Podaire, A., Bricaud, A., Buriez, J.-C., and Sèze, G.: The POLDER mission: Instrument characteristics and scientific objectives, IEEE Trans. Geosci. Remote Sens., 32(3), 598-615, 1994.

Deuzé, J. L., Bréon, F. M., Devaux, C., Goloub, P., Herman, M., Lafrance, B., Maignan, F., Marchand, A., Nadal, F., Perry, G., and Tanré, D.: Remote sensing of aerosols over land surfaces from POLDER-ADEOS–1 polarized measurements, J. Geophys. Res., 106, 4913–4926, 2001.

Diner, D. J., Bruegge, C. J., Martonchik, J. V., Ackerman, T. P., Davies, R., Gerstl, S. A. W., Gordon, H. R., Sellers, P. J., Clark, J., Daniels, J. A., Danielson, E. D., Duval, V. G., Klassen, K. P., Lilienthal, G. W., Nakamoto, D. I., Pagano, R., and Reilly, T. H.: MISR: A Multi-angle Imaging SpectroRadiometer for geophysical and climatological research from Eos. IEEE Trans. Geoscience and Remote Sens., 27 (2), 200-214, 1989.

Diner, D. J., Abdou, W. A., Ackerman, T. P., Crean, K., Gordon, H. R., Kahn, R. A., Martonchik, J. V., McMuldroch, S., Paradise, S. R., Pinty, B., Verstraete, M. M., Wang, M., and West, R. A.: Multi-angle Imaging SpectroRadiometer Level 2 aerosol retrieval algorithm theoretical basis, Revision G, JPL D-11400, Jet Propulsion Laboratory, California Institute of Technology, 2008.

~~European Space Agency, EnviSat AATSR product handbook, issue 2.2, 2007.~~

Foote, L.: Threshold considerations and wetland reclamation in Alberta's mineable oil sands, Ecology and Society, 17(1), 35, 2012.

Hillger, D., Kopp, T., Lee, T., Lindsey, D., Seaman, C., Miller, S., Solbrig, J., Kidder, S., Bachmeier, S., Jasmin, T., and Rink, T.: First-light imagery from Suomi NPP VIIRS, Bull. Amer. Meteor. Soc., 94, 1019-1029, 2013.

Holben, B., Eck, T., Slutsker, I., Tanre, D., Buis, J. P., Setzer, A., Vermote, E., Reagan, J. A., Kaufman, Y. J., Nakajima, T., Lavenu, F., Jankowiak, I., and Smirnov, A.: AERONET – A federated instrument network and data archive for aerosol characterization, Remote Sens. Environ., 66, 1–16, 1998.

Hsu, N. C., Jeong, M.-J., Bettenhausen, C., Sayer, A. M., Hansell, R., Seftor, C. S., Huang, J., and Tsay, S.-C.: Enhanced Deep Blue aerosol retrieval algorithm: The second generation, J. Geophys. Res. Atmos., 118, 9296–9315, doi:10.1002/jgrd.50712, 2013.

Hsu, Y.-M., Wang, X., Chow, J. C., Watson, J. G., and Percy, K. E.: Collocated comparisons of continuous and filter-based $PM_{2.5}$ measurements at Fort McMurray, Alberta, Canada, J. Air Waste Manage. Assoc., 66, 329-339, 2016.

Hubanks, P.: MODIS atmosphere QA plan for Collection 006, Greenbelt, MD, USA, NASA Goddard Space Flight Center, version 8, 2015.

Jiang, X., Liu, Y., Yu, B., and Jiang, M.: Comparison of MISR aerosol optical thickness with AERONET measurements in Beijing metropolitan area, Remote Sens. Environ., 107, 45–53, 2007.

Kahn, R. A., Gaitley, B. J., Martonchik, J. V., Diner, D. J., Crean, K. A., and Holben, B.: Multiangle Imaging Spectroradiometer (MISR) global aerosol optical depth validation based on 2 years of coincident Aerosol Robotic Network (AERONET) observations, J. Geophys. Res., 110, D10S04, doi:10.1029/2004JD004706, 2005.

Kahn, R. A., Gaitley, B. J., Garay, M. J., Diner, D. J., Eck, T. F., Smirnov, A., and Holben, B. N.: Multiangle Imaging SpectroRadiometer global aerosol product assessment by comparison with the Aerosol Robotic Network, J. Geophys. Res., 115, D23209, doi:10.1029/2010JD014601, 2010.

Kaufman, Y. J., Tanré, D., and Boucher, O.: A satellite view of aerosols in the climate system, Nature, 419, 215-223, 2002.

Kelly, F. J., Fuller, G. W., Walton, H. A., and Fussel, J. C.: Monitoring air pollution: Use of early warning systems for public health, Respirology, 17, 7-19, 2012.

Levy, R. C., Mattoo, S., Munchak, L. A., Remer, L. A., Sayer, A. M., Patadia, F., and Hsu, N. C.: The Collection 6 MODIS aerosol products over land and ocean, Atmos. Meas. Tech., 6, 2989–3034, 2013.

Levy, R. C., Munchak, L. A., Mattoo, S., Patadia, F., Remer, L. A., and Holz, R. E.: Towards a long-term global aerosol optical depth record: applying a consistent aerosol retrieval algorithm to MODIS and VIIRS-observed reflectance, Atmos. Meas. Tech., 8, 4083–4110, 2015.

Li, R.-R., Remer, L., Kaufman, Y. J., Mattoo, S., Gao, B.-C., and Vermote, E.: Snow and ice mask for the MODIS aerosol products, IEEE Geosci. Remote Sens. Lett., 2, 306-310, 2005.

Li, Z., Zhao, X., Kahn, R., Mishchenko, M., Remer, L., Lee, K.-H., Wang, M., Laszlo, I., Nakajima, T., and Maring, H.: Uncertainties in satellite remote sensing of aerosols and impact on monitoring its long-term trend: a review and perspective, Ann. Geophys., 27, 2755–2770, 2009.

Li, C., Hsu, N. C., Sayer, A. M., Krotkov, N. A., Fu, J. S., Lamsal, L. N., Lee, J., Tsay, S.-C.: Satellite observation of pollutant emissions from gas flaring activities near the Arctic, Atmos. Environ., 133, 1-11, 2016.

Liggio, J., Li, S.-M., Hayden, K., Taha, Y. M., Stroud, C., Darlington, A., Drollette, B. D., Gordon, M., Lee, P., Liu, P., Leithead, A., Moussa, S. G., Wang, D., O'Brien, J., Mittermeier, R. L., Brook, J., Lu, G., Staebler, R., Han, Y., Tokarek, T. T., Osthoff, H. D., Makar, P. A., Zhang, J., Plata, D., Gentner, D. R.: Oil sands operations are a major source of secondary organic aerosols, Nature, 534, 91-94, 2016.

Liu, Y., Sarnat, J. A., Coull, B. A., Koutrakis, P., and Jacob, D. J.: Validation of Multiangle Imaging Spectroradiometer (MISR) aerosol optical thickness measurements using Aerosol Robotic Network (AERONET) observations over the contiguous United States, J. Geophys. Res., 109, D06205, doi:10.1029/2003JD003981, 2004.

Loría-Salazar, S. M., Holmes, H. A., Arnott, W. P., Barnard, J. C., Moosmüller, H.: Evaluation of MODIS columnar aerosol retrievals using AERONET in semi-arid Nevada and California, U.S.A., during the summer of 2012, Atmos. Environ., 144, 345-360, 2016.

Lyapustin, A., Wang, Y., Laszlo, I., Kahn, R., Korkin, S., Remer, L., Levy, R., and Reid, J. S.: Multiangle implementation of atmospheric correction (MAIAC): 2. Aerosol algorithm, J. Geophys. Res., 116, D03211, doi:10.1029/2010JD014986, 2011.

McLinden, C. A., Fioletov, V., Boersma, K. F., Krotov, N., Sioris, C. E., Veefkind, P., and Yang, K.: Air quality over the Canadian oil sands: A first assessment using satellite observations, Geophys. Res. Lett., 39, L04804, http://dx.doi.org/10.1029/2011GL050273, 2012.

McLinden, C. A., Fioletov, V., Krotkov, N., Li, C., Boersma, K. F., and Adams, C.: A decade of change in NO_2 and SO_2 over the Canadian oil sands as seen from space, Env. Sci. Tech., doi:10.1021/acs.est.5b04985, 2016.

Myhre, G., Shindell, D., Bréon, F.-M., Collins, W., Fuglestvedt, J., Huang, J., Koch, D., Lamarque, J.-F., Lee, D., Mendoza, B., Nakajima, T., Robock, A., Stephens, G., Takemura, T., and Zhang, H.: Anthropogenic and Natural Radiative Forcing. In: Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change [Stocker, T.F., D. Qin, G.-K. Plattner, M. Tignor, S.K. Allen, J. Boschung, A. Nauels, Y. Xia, V. Bex and P.M. Midgley (eds.)]. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, 2013.

O'Neill, N. T., Eck, T. F., Holben, B. N., Smirnov, A., Royer, A., and Li, Z.: Optical properties of boreal forest fire smoke derived from Sun photometry, J. Geophys. Res., 107, 4125, doi:10.1029/2001JD000877, 2002.

Popp, T., de Leeuw, G., Bingen, C., Brühl, C., Capelle, V., Chedin, A., Clarisse, L., Dubovik, O., Grainger, R., Griesfeller, J., Heckel, A., Kinne, S., Klüser, L., Kosmale, M., Kolmonen, P., Lelli, L., Litvinov, P., Mei, L., North, P., Pinnock, S., Povey, A., Robert, C., Schulz, M., Sogacheva, L., Stebel, K., Zweers, D. S., Thomas, G., Tilstra, S. Vandenbussche, L. G., Veefkind, P., Vountas, M. and Xue, Y.: Development, production and evaluation of aerosol Climate Data Records from European satellite observations (Aerosol_cci), Remote Sensing, 8, 421, doi:10.3390/rs8050421, 2016.

Sayer, A. M., N. C. Hsu, C. Bettenhausen, and M.-J. Jeong (2013), Validation and uncertainty estimates for MODIS Collection 6 "Deep Blue" aerosol data, J. Geophys. Res. Atmos., 118, 7864–7872, doi:10.1002/jgrd.50600.

Stieb, D. M., Burnett, R. T., Smith-Doiron, M., Brion, O., Shin, H. H., and Economou, V.: A New Multipollutant, No-Threshold Air Quality Health Index Based on Short-Term Associations Observed in Daily Time-Series Analyses, J. Air & Waste Manage. Assoc., 58(3), 435-450, 2008.

Tian, J., and Chen, D.: Spectral, spatial, and temporal sensitivity of correlating MODIS aerosol optical depth with ground-based fine particulate matter ($PM_{2.5}$) across southern Ontario, Can. J. Remote Sensing, 36, 119–128, 2010.

van Donkelaar, A., Martin, R. V., Brauer, M., Hsu, N. C., Kahn, R. A., Levy, R. C., Lyapustin, A., Sayer, A. M. and Winker, D. M.: Global estimates of fine particulate matter using a combined geophysical-statistical method with information from satellites, models, and monitors, Environ. Sci. Technol., 50, 3762−3772, 2016.

Wu, A., Xiong, X., ~~Sun,~~Doelling, D. R., Morstad, D., Angal, A~~Esposito,Guenther, B~~., and ~~Barnes, W.;~~Bhatt, R.: Characterization of Terra and Aqua MODIS ~~reflective solar bands~~VIS, NIR, and SWIR spectral bands' calibration ~~algorithm and on-orbit performance, Proc. SPIE, 4891, 95-104, 2003.~~ stability, IEEE Trans. Geosci. Remote Sens., 51, 4330-4338, 2013.

Appendix A: Data product notes

MODIS data is obtained from ftp://ladsweb.nascom.nasa.gov/allData/. AATSR and ATSR-2 version 4.1 data are from Swansea University and can be obtained from the Aerosol CCI website (http://www.esa-aerosol-cci.org/) following registration. The current file version (F12) is used for MISR (ftp://l5eil01.larc.nasa.gov/MISR/MIL2ASAE.002). The selected MISR AOD product is named the "regional best estimate of spectral optical depth". POLDER data was obtained from CNES (http://polder.cnes.fr), but data can currently be obtained from http://www.icare.univ-lille1.fr/ following registration. A POLDER AOD datum is filtered if any of the following statements are true (see ~~F. M.~~ Bréon~~,~~ (2011~~);~~)):

1) The central pixel is snow-covered.
2) One of the cloud tests is not applied.
3) None of the 9 radiance pixels which form the AOD superpixel has clear sky.
4) Sufficient data couples do not exist. The couples are:
   a) 865 nm & 910 nm,
   b) Q443 & U443,
   c) Q670 & U670,
   d) Q865 & U865,

   where Q and U are the derived Stokes elements and the number is the wavelength (in nm) of the channel.
5) Ozone absorption is not corrected (using TOMS or ECMWF).
6) Stratospheric aerosol correction is uncertain or imprecise (i.e. stratospheric AOD larger than a certain threshold).
7) Minimum scattering angle is larger than a threshold or maximum scattering angle is smaller than a threshold.
8) Aerosol optical thickness is larger than a threshold such that surface reflectance cannot be estimated adequately.
9) A large difference between measured and modeled reflectance exists for 443 nm.
10) Differences are too large between measured and modeled reflectance (risk of glitter).
11) Meteorological data indicate the presence of snow at ground level.
12) The quality index is 0.00 for viewing geometry conditions
13) The quality index is 0.00 for polarized reflectance fit.

Satellite	Time period	Wavelength (nm)	Spatial resolution of AOD superpixel at nadir (km^2)	Spatial resolution of radiances (km^2)
MISR	2000-2015	558	17.6 × 17.6	1.1 × 1.1
MODIS: Terra	2000-2015	470, 550, 660	10 × 10 (also 3 × 3)	0.25 × 0.25 to 1~~×~~ × 1
Aqua	2002-2015			
POLDER: 1	1996-1997	865	18 × 21	6 × 7
2	2003			
(PARASOL) 3	2005-2013			
ATSR: ATSR-2	1995-2003	550	10 × 10	1 × 1
AATSR	2002-2012			

Table 1. Spatial resolution of AOD data products from selected satellite instruments. The third column contains the wavelength at which aerosol optical depth is reported in each satellite data product. MISR and both MODIS instruments are currently operating.

Station name	lat(°N)	lon(°W)	Time span
Anzac	56.4493	-111.0372	2006-2015
Fort McMurray Athabasca Valley	56.7328	-111.39	1997-2015
Fort McMurray Patricia McInnes	56.7522	-111.476	1999-2015
Millenniummine	56.97	-111.4	2001-2015
Syncrude Upgrader Expansion 1	57.1492	-111.642	2002-2015
Fort McKay	57.1894	-111.641	1997-2015
Wapasu	57.2383	-110.9028	2013-2015
Shell Muskeg River	57.2491	-111.508567	2009-2015
Albian mine	57.2808	-111.526	2001-2009
Canadian Natural Resources Ltd. Horizon	57.3037	-111.739617	2008-2015

Table 2. Selected NAPS $PM_{2.5}$ sites and time span of available data (inclusive)

AOD product	R	N
POLDER/PARASOL 865 nm	0.~~83~~64	8
AATSR 550 nm	0.~~77~~73	9
MISR 558 nm	~~-~~0.~~41~~20	10
MODIS/Aqua DT ~~470~~550 nm	0.~~49~~23	10
MODIS/Aqua DB 550 nm	0.~~81~~57	10

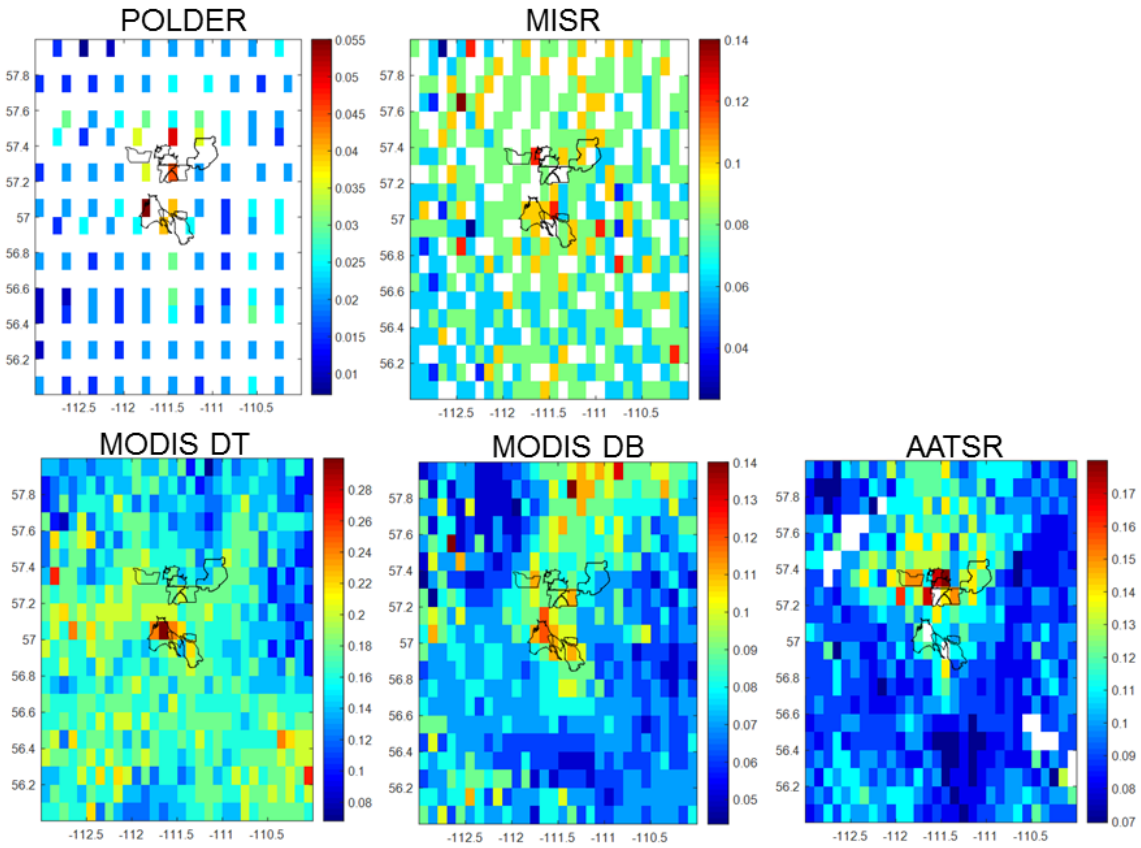
Table 3. Spatial correlation between $PM_{2.5}$ mass density and AOD using ~~means~~medians of coincident data ~~over the entire overlapping period~~ at 10 sites in the AOSR. Wapasu has insufficient or no temporal overlap with POLDER/PARASOL and AATSR. Syncrude UE1 is not spatially coincident with any of the POLDER locations given the 10 km criterion (see Sect. 2).

sensor	~~r~~ r_s	slope	offset	seasonal *r*	month range	N
Aqua DB v6	0.~~81~~793 0.~~94~~686	0.~~89~~867 1.~~00~~013	0.~~0304~~0233 0.~~0171~~0036	0.84 0.84	4-10 4-10	5508 626
Aqua DT v6	0.~~956~~875 0.~~972~~806	1.~~11~~210 1.~~08~~088	-0.~~0013~~0117 -0.~~0177~~0223	0.99 0.~~959~~95	4-10 5-9	~~4748~~ ~~4084~~4197 431
PARASOL	0.~~92~~867 -	~~1.09~~ 0.831 -	-0.~~03~~018 -	0.89 -	5-10 -	414 -
AATSR	0.~~91~~888 -	0.~~88~~862 -	0.~~0265~~0253 -	0.96 -	5-10 -	560 -
MISR	0.~~89~~862 0.~~93~~668	0.~~63~~731 0.~~64~~652	0.~~0293~~0219 0.~~0364~~032	0.88 -	3-9 -	337 87

Table 4. Statistical comparison of coincident AODs observed by satellite-based sensors and AERONET ~~CIMEL~~Cimel sun ~~photometer~~photometers. For each satellite AOD product, the upper row is for Fort McMurray and the lower row is for Fort McKay. The ~~CIMEL~~Cimel 500 nm AOD, scaled to the satellite AOD wavelength (see Sect. 2), is used for comparison with all satellite sensors except POLDER/PARASOL, for which the ~~CIMEL~~Cimel 870 nm AOD is more appropriate (see Table 1). ~~The simple linear regression equation used to obtain the slope and offset assumes AERONET AOD and satellite-based AOD are the independent and dependent variables, respectively.~~ The number of MISR-Fort McKay coincidences is insufficient to assess the month-to-month variability.

	Including +4σ outliers	Excluding +4σ outliers
POLDER/PARASOL	0.995	0.81
MISR	0.91	0.94
AATSR	0.98	0.92
MODIS DT	0.~~97~~98	0.~~95~~94
MODIS DB	0.91	0.86

Table 5. Correlation of annual mean AODs with Fort McMurray AERONET AODs during the respective overlap periods of the various satellite AOD products. In the rightmost column, the contribution of large forest fires has been removed from AERONET data and satellite datasets using +4 standard deviations (σ) as a cutoff.


POLDER
MISR
MODIS DT
MODIS DB
AATSR

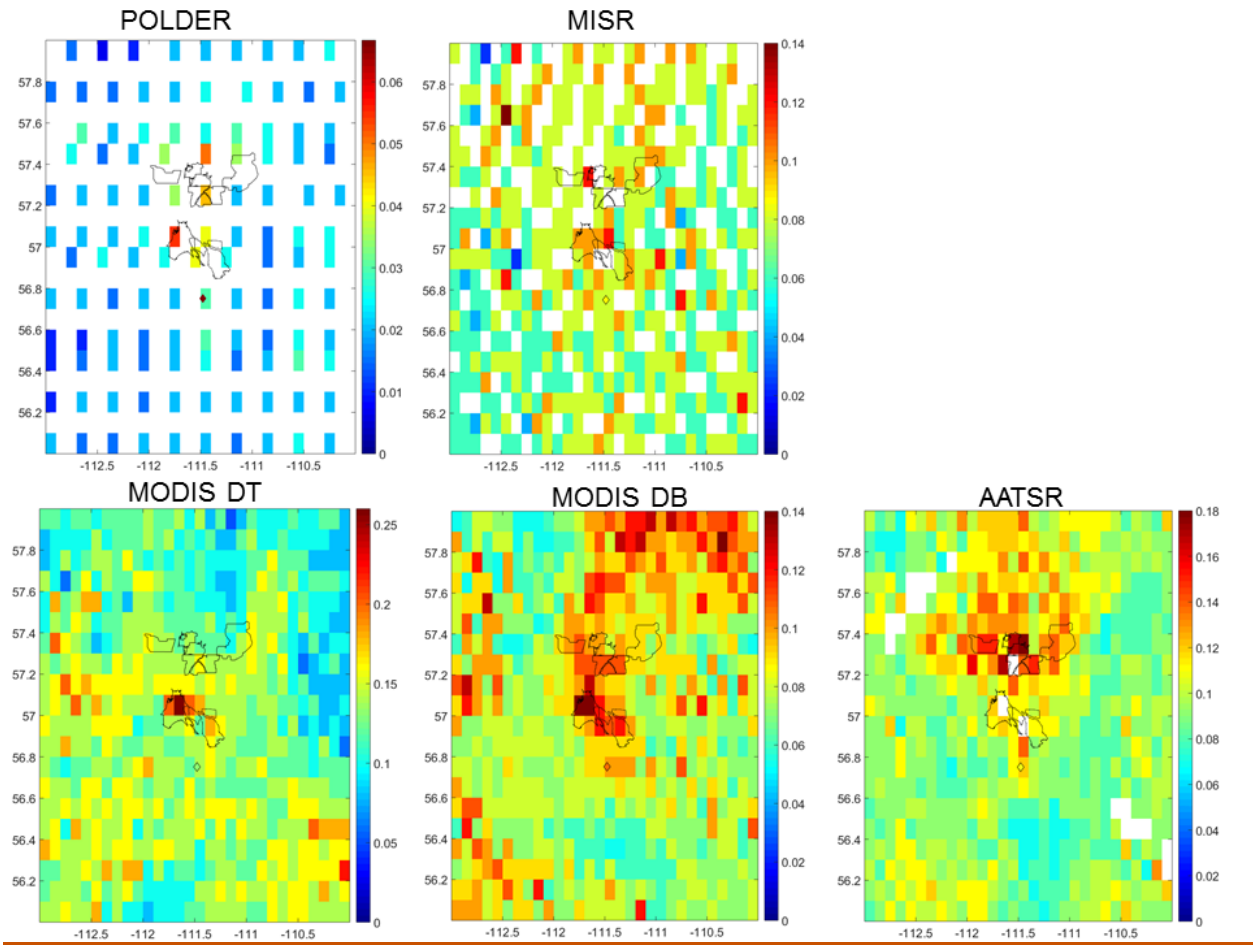


Figure 1. Climatological average AOD maps on a 0.1° x 0.1° latitude-longitude grid. (top left) POLDER 865 nm (1996-2013). Note the gaps in time between the different members of the POLDER series in Table 1. (top right) MISR 558 nm (2000-2015). (bottom left) MODIS/Aqua DT using only confidence of 3 (2002-2015). (bottom centre) MODIS/Aqua DB using only ~~confidence~~confidences of 2-3 (2002-2015~~).~~) following Sayer et al., 2013. (bottom right) AATSR 550 nm (2002-2012). Typical N is ~65 for AATSR (see below) and white areas indicate N<20. Black lines trace out the three surface mining areas in this and subsequent figures. —Average coincident AERONET AOD at Fort McMurray is superimposed as a diamond with a black outline. In each panel, the AOD ranges from 0 to the greater of the maximum climatological mean satellite AOD or the Fort McMurray AERONET mean AOD, except for MISR, for which the AOD range is capped at 0.14 to not emphasize the anomalously high AOD at Moose Lake (57.6°N, 112.5°W).

1

2

3

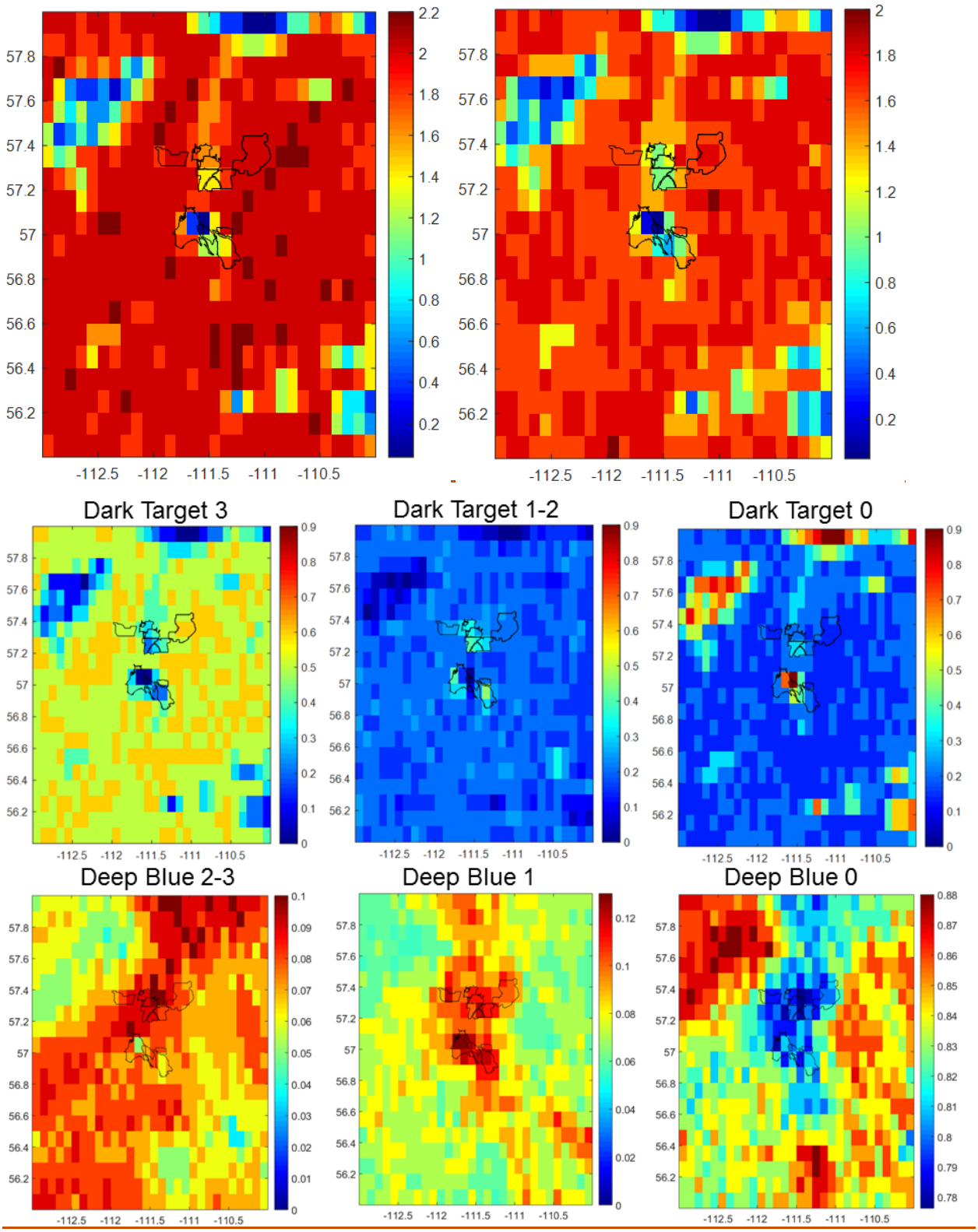


Figure 2. ~~Map~~Maps of ~~climatological average~~fraction of pixels with a specific confidence (2002-2014) for MODIS/Aqua DT (~~left~~top) and DB (~~right~~bottom) AODs. Lower confidence is expected over Moose Lake ~~(57.6°N,~~

~~112.5°W)~~ and the Richardson sand dunes (58.0°N, 111.0°W). MODIS DB only reports a fill value when confidence is 0 in contrast to MODIS DT, thus the bottom right plot accounts for fill values, whereas the top right plot (for MODIS DT) does not.

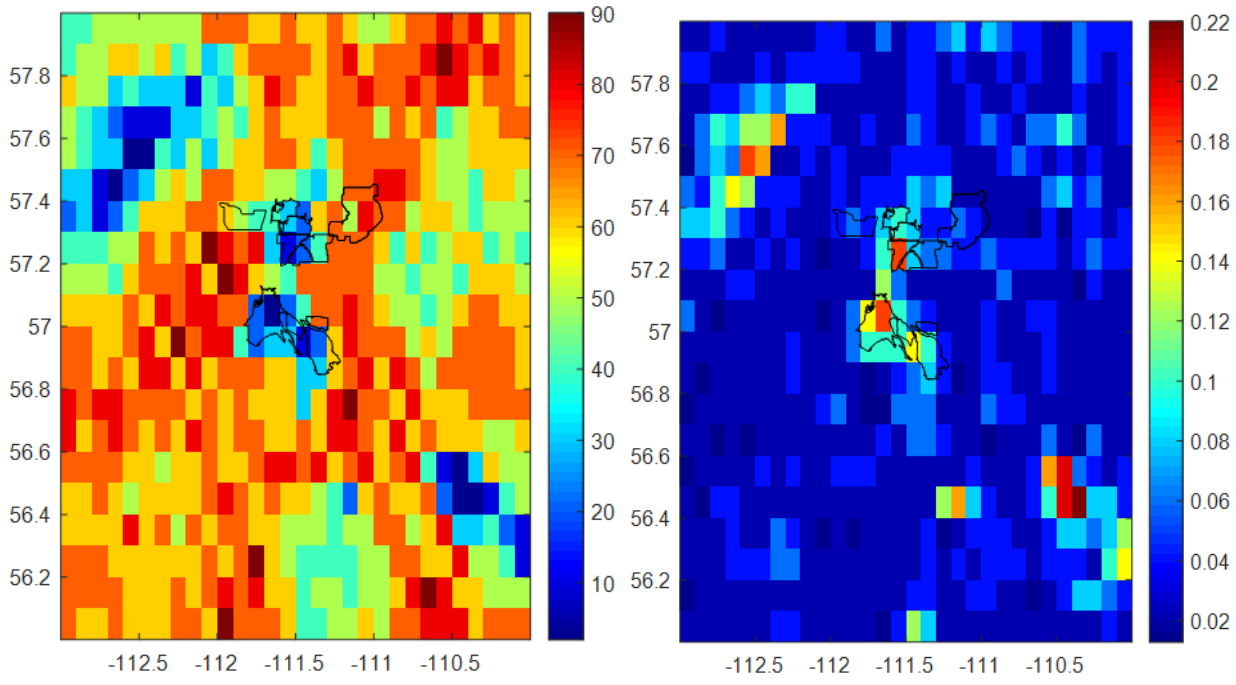


Figure 3. Map of sample size (left) and average cloud fraction within AOD superpixels when the AOD retrieval is successful (right), compiled from the entire AATSR data record. Smaller sample sizes are expected over Moose Lake and Gordon Lake (56.5°N, 110.5°W).

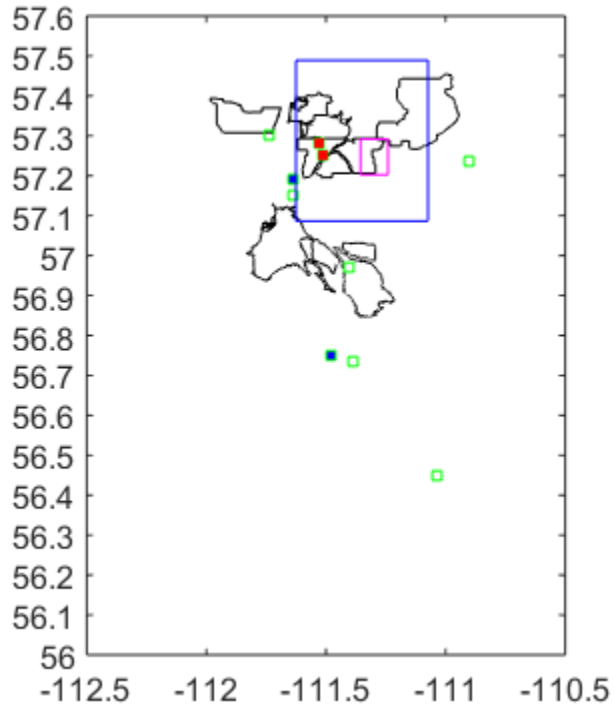
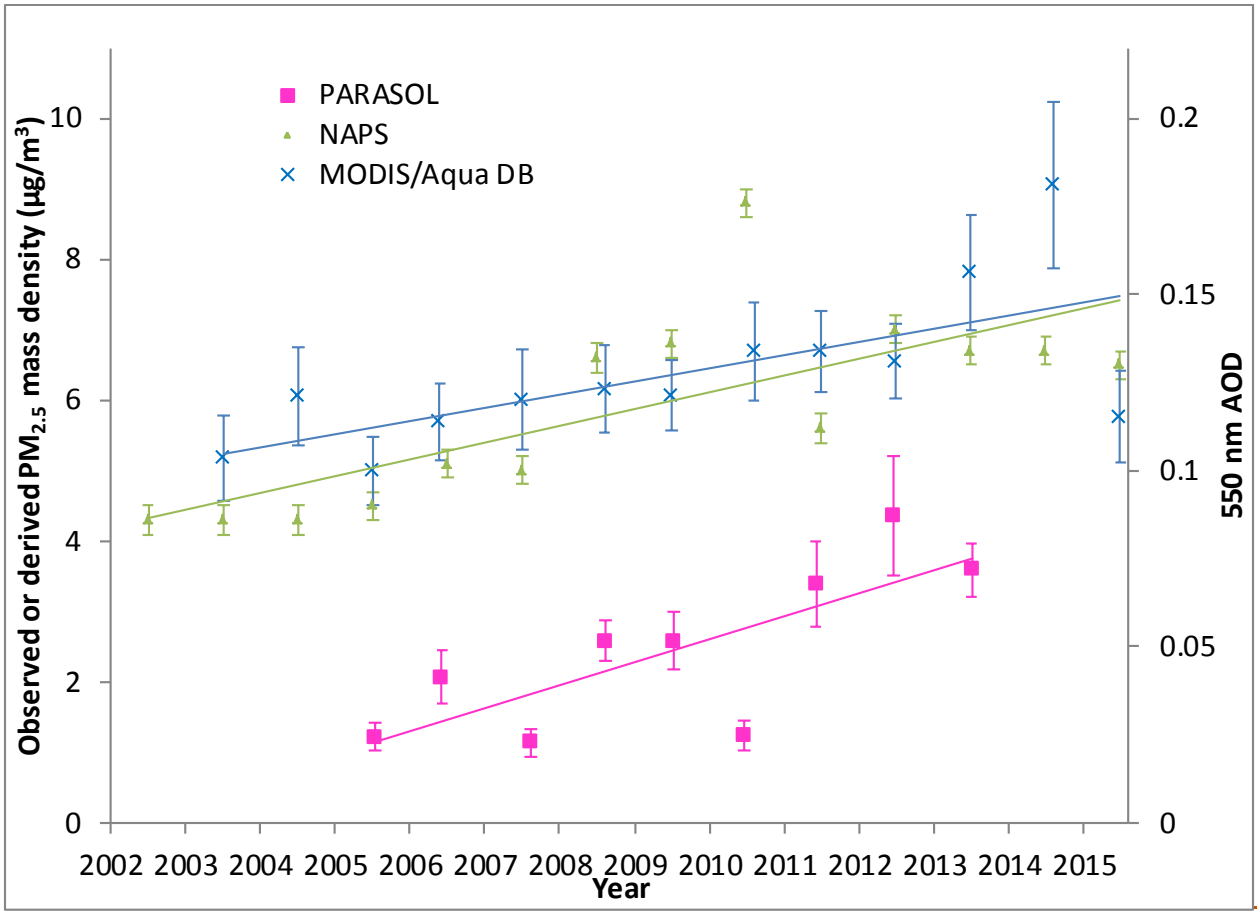


Figure 4. Areas with a significant positive trend in AOD in the POLDER/PARASOL, and MODIS/Aqua DB data records. The area over which the AOD time series is determined for MODIS/Aqua DB (0.1×0.1°), and POLDER/PARASOL (0.4×0.4°) is outlined in pink and blue, respectively. Locations of 10 NAPS $PM_{2.5}$ monitoring sites are also shown as small green squares. The central one of 3 adjacent (overlapping) grid cells at constant latitude is plotted for POLDER/PARASOL (see Sect. 3 for details). The grid cell with the largest trend in the domain is plotted for MODIS/Aqua DB (see Sect. 3 for details). Note that the Albian mine site (57.2808°N, 111.526°W) was replaced by the nearby Shell Muskeg River site (57.2491°N, 111.509°W) in 2009 (both station symbols are filled in red). The two AERONET instruments are co-located with NAPS monitors and those sites are filled in blue.

1


PARASOL
NAPS
MODIS/Aqua DB
Observed or derived PM2.5 mass density (µg/m3)
550 nm AOD
Year
0
2
4
6
8
10
0
0.05
0.1
0.15
0.2
2002 2003 2004 2005 2006 2007 2008 2009 2010 2011 2012 2013 2014 2015

2

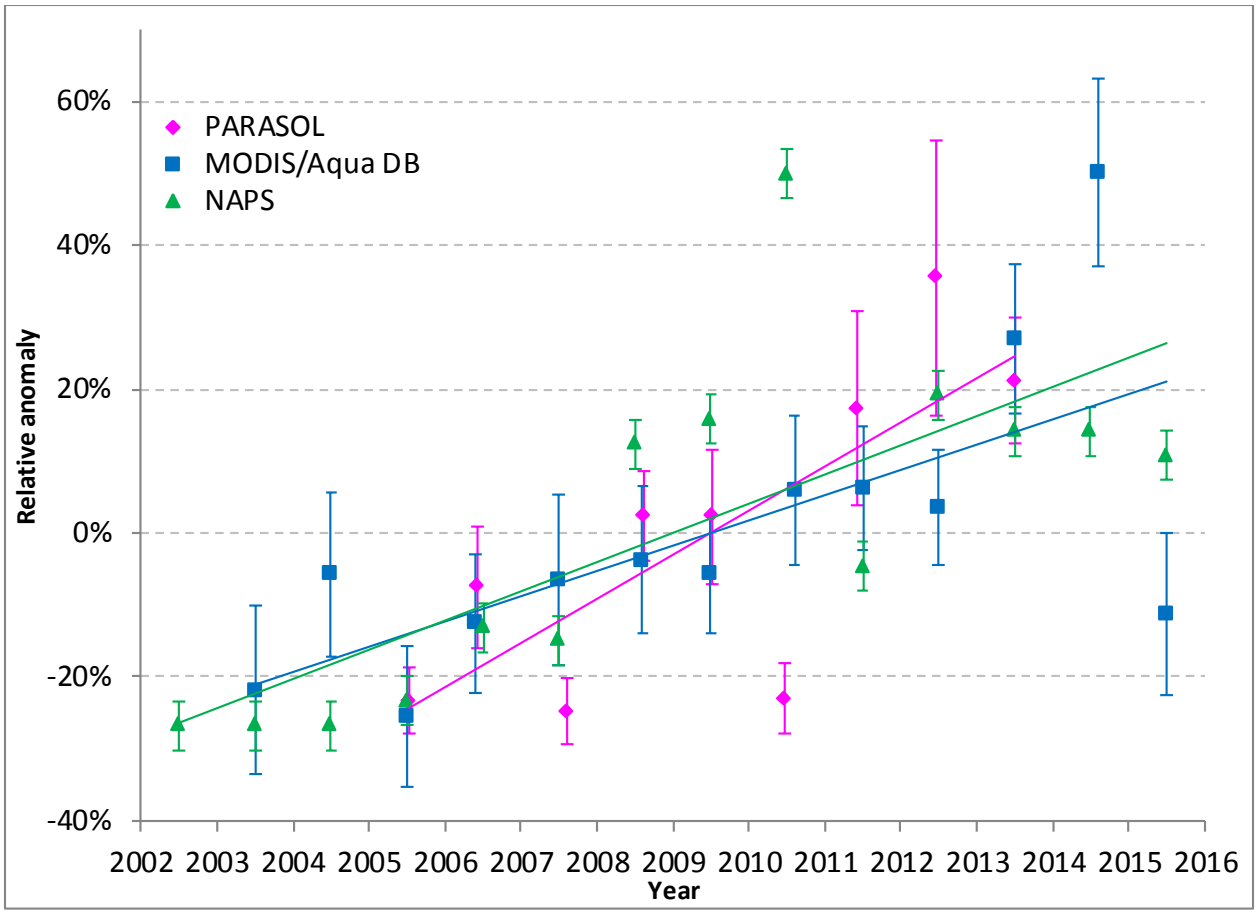
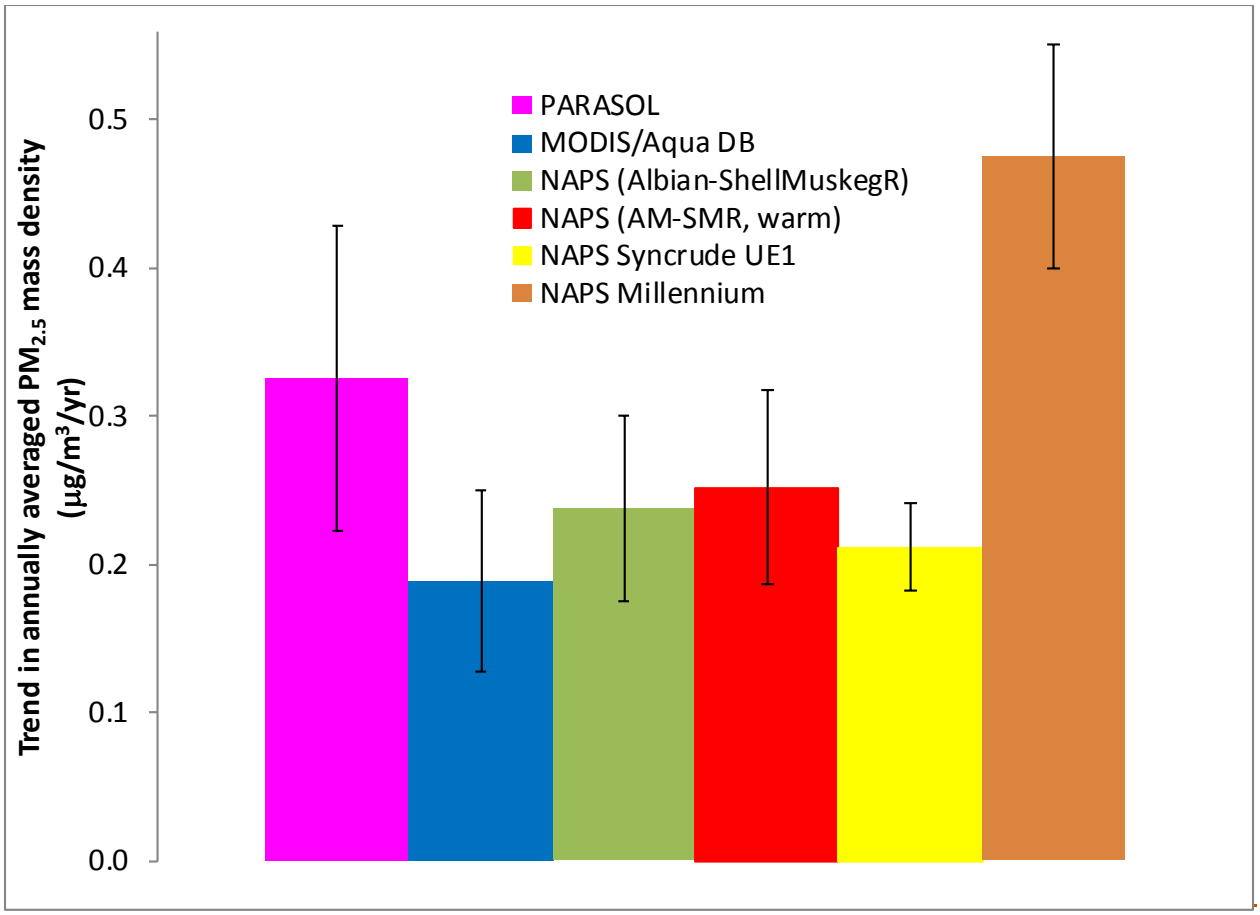


Figure 5. ~~Annual average~~Relative anomalies of annual mean $PM_{2.5}$ mass density for the merged Albian mine and Shell Muskeg River dataset, along with ~~$PM_{2.5}$ annual averages derived from satellite AOD data records~~relative anomalies of bias-corrected annual mean AODs for POLDER/PARASOL and MODIS/Aqua DB (see Sect. 3 for details and Fig. 4 for satellite trend areas). Each satellite time series is plotted at the average decimal time for each calendar year. Trend lines are fitted to each time series using a matching colour. Vertical error bars indicate ±1 relative standard error of the annual mean. There are, on average, 33 and 50 observations per year for POLDER/PARASOL and MODIS/Aqua DB, respectively. ~~The secondary ordinate applies to the MODIS DB observations, but not POLDER/PARASOL (for which the 865 nm AODs are in the 0.01 to 0.03 range).~~


Trend in annually averaged $PM_{2.5}$ mass density ($\mu g/m^3/yr$)
0.0
0.1
0.2
0.3
0.4
0.5
PARASOL
MODIS/Aqua DB
NAPS (Albian-ShellMuskegR)
NAPS (AM-SMR, warm)
NAPS Syncrude UE1
NAPS Millennium

1

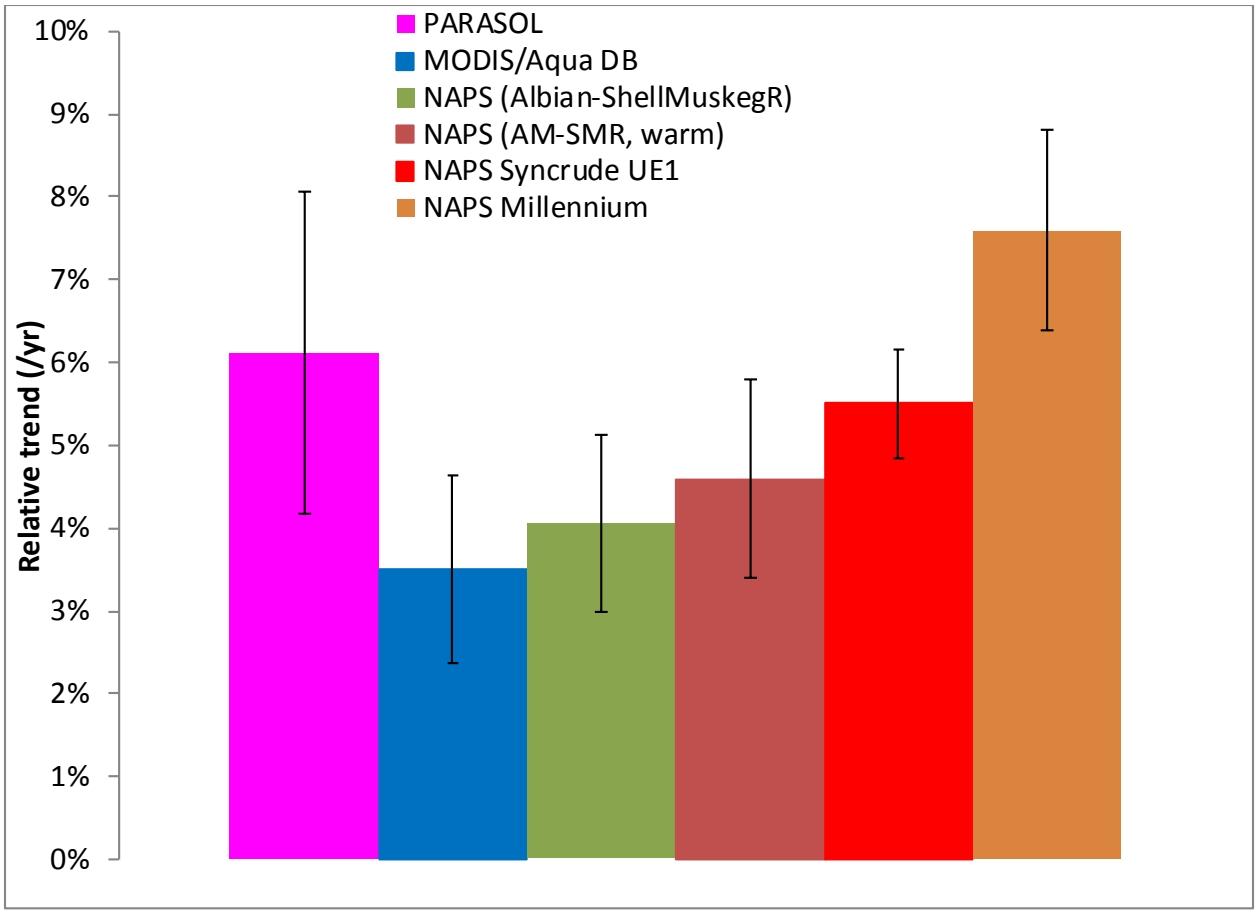


Figure 6. ~~Trend~~Relative trend in annually averaged $PM_{2.5}$ mass density calculated using NAPS $PM_{2.5}$ data for three locations, namely the merged Albian mine and Shell Muskeg River dataset (2002-2015), Millennium mine (2002-2014) and Syncrude UE1 (2003-2014), or ~~derived from~~in satellite AODs in the vicinity of Shell's Albian and Muskeg River mines (see Fig. 4 and Sect. 3). The trend is also determined for the NAPS $PM_{2.5}$ merged Albian Mine – Shell Muskeg River (AM-SMR) dataset limiting to the warm season (April to October). Trend uncertainty is indicated with a vertical bar (±1 s. e.).