Dear Editor, dear Dr. Andrew Sayer,

Many thanks for the valuable comments, which help to improve the quality of this paper. The detailed replies are addressed below point by point.

As below, I would like to clarify some of the points raised by the reviewers. We hope that the reviewers and the editors will be satisfied with our responses to the 'comments' and the revisions for the original manuscript.

Best regards,

Yahui Che on behalf of all authors

2019-5-15

Response to reviewer 1:

Interactive comment on "Investigations into the Development of a Satellite-Based Aerosol Climate Data Record using ATSR-2, AATSR and AVHRR data" by Yahui Che et al.

Andrew Sayer (Referee)
andrew.sayer@nasa.gov

Received and published: 20 February 2019

I am writing this review under my own name (Andrew Sayer) as I have worked with all the satellite instruments the authors use here (although on different algorithms for the ATSRs/AVHRRs), am one of the people who developed the MODIS C6.1 aerosol products used, and am familiar with the authors' work on this topic from their previous studies and discussions at conferences.

This paper seeks to assess whether the gap in aerosol optical depth (AOD) data between the end of the AATSR record in 2012 and start of the SLSTR mission in 2016 (using an algorithm developed by co-author de Leeuw's group) can be bridged with an AVHRR-based algorithm developed by the other authors. Since we are now at the stage where we have multiple satellite records of decent length and quality, it is reasonable to ask how these might be best combined and what sort of consistency can be achieved. So on that front the study is relevant and important. The authors focus on north-eastern China, and in particular two AERONET sites in the Beijing area.

In all honesty the choice of AVHRR and the Beijing region to attempt to bridge the AATSR/SLSTR records seems to be guided in part by the authors' own created data sets and home institution, as the AVHRR sensor is perhaps the least capable of satellite instruments which could have been chosen to fill in this gap, and aerosols in China are complicated so perhaps not the place in the world you'd want to start developing a framework for combining records. The concepts behind the ATSR and AVHRR AOD retrieval algorithms used, and their sampling limitations, are also quite different, which would also affect the expected level of consistency. For example, you might expect a MISR-based algorithm to be a better bridge because MISR and the ATSRs are all multiangle, narrow-swath sensors. I suppose I am left asking: why did the authors try to answer question "how well can I bridge the AOD gap between AATSR and SLSTR in north-eastern China using this AVHRR data set?" rather than the more general questions "how well do available satellite records allow us to create a long-term AOD record over north-eastern China?" or "how should we combine different satellite aerosol records?". This analysis is a step in that direction which can provide input to answering those general questions, so I think it does have value. I can understand some scientific rationale behind this because as the authors point out the AVHRRs are the longest available sensor series (going back to around 1980), and that part of the world has undergone a lot of change in that period with considerable aerosol sources. So scientifically it makes some sense to look at this region and sensor. However I would have liked to see that aspect emphasised more in the paper: the title implies a bigger scope, and the bulk of the analysis in the paper sticks to the post-2000 era.

My overall recommendation is for major revisions; I would like to review the revised version. There is value in this analysis but I think some things need clarification, some need extention, and more

big-picture guidance about the insights the authors have got would be useful for readers planning follow-on studies on this topic. Some specific comments and suggestions relating to these general comments are below:
Re: Thank you for your recommendation and give us a chance to improve the manuscript. We have addressed these points in the revised manuscript.

Title: I suggest adding "over north-eastern China" or similar after "Climate Data Record", to better fit the scope and content of the paper. The current title implies a more large-scale analysis.
Re: We revised the title by adding "over North-Eastern China from 1987 to 2012" to indicate both time and regions.

Introduction: SLSTR was launched in 2016 and this paper is largely about bridging the 2012-2016 gap. However, no SLSTR data are shown in the study. Somewhere in here, could the authors add a few statements about the current status of SLSTR AOD data? If there is a SLSTR product in development, it would make sense to wait until that can be included in the analysis (even if it is preliminary). That way we could answer the question of whether AVHRR can bridge the gap, by seeing how consistent the results are with actual SLSTR retrievals. If there is no SLSTR product in development then ok, but this should be stated, and the premise of the paper becomes more questionable (why try to bridge if there is nothing at the other end of the bridge?). On a related note, the AVHRR data processed here end in 2014 (although the AVHRRs are still flying) so would not be able to bridge the gap anyway unless the authors extend the processing period.
Re: Thanks for this remind. To the current, SLSTR data are not available. Hence, we switched to extend ATSR data set back to 1980s, no longer focus on bridging gap between AATSR and SLSTR.

Introduction: As noted in my general comments, I'd like to see more discussion of the bigger picture of why we want to combine satellite records, how it has been done, what the challenges are, and why these sensors and this region were chosen as an example.
Re: Referee 3 also mentioned this point. We switched focus on making an extention back to 1980s for ATSR data set, no longer pay attention to bridge the gap of AATSR and SLSTR. Corresponding revisions have been made in new version of manuscript.

Page 6 line 4: I suggest citing a paper here (there are lots of AVHRR aerosol/cloud/surface papers which give instrument descriptions) rather than an ESA webpage, especially for a non-ESA sensor.
Re: Here we update the citation to "NOAA KLM user's guide" from NOAA official website.

Page 6, lines 15-25: Surface reflectance modelling is one of the largest error sources for AOD retrieval, especially for a sensor like AVHRR which has only a few, and quite broad, spectral bands. The authors' algorithm relates 3.75 micron to 0.64 micron reflectance using an empirical relationship which is a function of surface type. The surface cover type used is obtained from the MODIS MCD12C1 product. This is static for a given location, therefore, if the surface cover changes during the period, the AOD retrieval error will change through time. This is not discussed in the paper but is quite important if the goal is to make a data record going back to the 1980s, as we know there has been a lot of industrial station since then. Basically we know that surface type has changed over parts of this region, so we know that this key algorithm assumption has been violated.

Yet I did not see this discussed in this paper, or in the previous algorithm papers cited here. There should be some discussion and some quantification of the effects of an incorrect surface type classification on the retrieved AOD. (For example, if you use the model for type A instead of type B, how much does the retrieved AOD change, and how systematic is that?) It is also not clear that this surface model can account for the (non-negligible) differences in the spectral response functions between the different sensors, which could result in systematically different reflectances observed for the same underlying surfaces. If these aren't addressed then it undercuts the idea of trying to get a long term data record.

Re: This version of AVHRR AOD data set does have some limitations or disadvantages. As you know, it is very difficult to retrieval AOD when there only one visible band, besides there is no enough prior knowledge in the early time. This paper focus on the analysis of the first version of AVHRR AOD results but not the algorithm itself. Now we are trying to make more detailed validation and improve this algorithm. However, the newly version of long term AVHRR AOD data set is not available now.

Page 6 line 30: NOAA15 says 199-2002, I guess this should be 1998 or 1999 rather than 199? Also, TETOP-A should read METOP-A.

Re: The data from NOAA-15 is from 1992 to 2002, so it should be 1999. These two typos have been revised in new version of manuscript.

Page 8 line 14 and later: The authors sometimes refer to "radiance-derived AOD", sometimes "solar radiation", and sometimes "broadband extinction", when describing one of the ground-based data sets. These phrases, so far as I can tell, all refer to the same data set, but they do not have any words in common. I suggest keeping the terminology consistent to make things clear. My personal preference would be to say "broadband solar radiation" but I'd leave that up to the authors. This refers to some figure captions as well as the body text.

Re: Here we use "BEM (broadband extinction method) AOD" throughout new manuscript, as it does in Dr. Ling Sun's paper. https://doi.org/10.1016/j.atmosenv.2015.08.042

Page 9 line 5: There is now an AERONET version 3 paper published, which could be cited here. https://www.atmos-meas-tech.net/12/169/2019/

Re: Thanks. We have citied this paper in new version of manuscript.

Page 10 line 21: This says that the AVHRR retrievals are available from 1983-2014. But from page 6, only data back to 1987 are used in this study. Why not include the first four years too, since the authors state they processed the data?

Re: All instruments we selected in this paper are onboard satellites along with descending orbit. For years from 1983 to 1986, AVHRR raw data are not consistent, data only from 1986.11 to 1986.12 (from NOAA-10), 1985.07 to 1985.10 (from NOAA-8), 1984.05 to 1984.06 (from NOAA-8) and 1984.07 (from NOAA-6), and 1983.01 to 1983.03 (from NOAA-6) are available. So we abandon data from 1983 to 1986. This is stated in new version of manuscript.

Figure 2: We know that AOD changes a lot in different seasons. We also know that satellite biases are affected by geometry, AOD/aerosol type, and surface reflectance, which also change a lot in

different seasons. We also know that sampling is related to cloud and snow cover, which also change a lot in different seasons. We know that all these factors can affect different algorithms in different ways. Given that, I do not find multiannual means as presented in this Figure to be useful, as we have no idea which factors are contributing to the differences observed here. I suggest changing this to show multiannual seasonal composites instead of overall multiannual means. I feel that will be more useful and allow for more insight about these confounding factors. Otherwise the authors need to provide a justification why presenting the data in this way is useful, given these issues.

Re: Thanks, seasonal characteristics of differences and spatial distribution are really helpful for understanding AOD retrieval algorithms and data sets. Seasonal AOD and AOD differences have been added into new version of manuscript.

Page 12, line 23: should be changed to "to MODIS Dark Target over land", for completeness. This expression does not apply to the Deep Blue or merged product (which is what is actually used here).

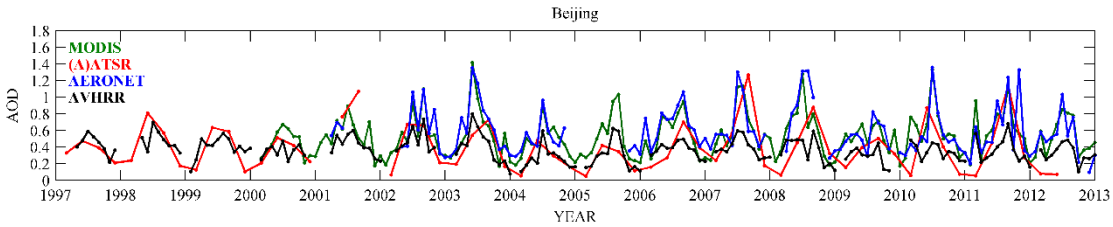
Re: This has been revised in new version of manuscript.

Page 13, lines 28-30: If my reading is correct, the authors are saying that the MODIS C6.1 validation results obtained here have different error characteristics than the MODIS C6.1 validation results over the same region obtained by Sogacheva et al (2018). Is that correct? If so, it seems surprising, so needs to be examined in more detail. Is there an error in one of the studies? Where does the difference come from? If it is a result of a different sampling period or sites, that is important to discuss, as it indicates that the regional validation is not robust and therefore that we cannot take these results as representative. This should be clarified in the text.

Re: Yes, that is correct, but the region is not e exactly the same. Moreover, the difference is that we used data from both AERONET and CARSNET in validation while Sogacheva et al (2018) only used AERONET data in validation. This is why there are different error characteristics. As shown by fig.1 in manuscript, all AERONET sites are in or around Beijing where main surface types are urban, croplands or mixed forest and aerosol sources are mainly anthropogenic, on the contrary CARSNET sites distribute in more regions in study area especially in sparsely vegetated regions where the surface and aerosol circumstances are significantly different. Hence, the conclusions are with small differences. This has been clarified in the new version of manuscript.

Figures 4-6 (and 9, 10): I find these a little hard to interpret. We are trying to see how consistent the data sets are with AERONET and each other, yet by plotting all of them on separate panels it makes harder to see how consistent they are. I suggest plotting all the data sets on a single panel instead, so we can directly compare them. I realize that the ATSR data shown are seasonal while the others are monthly, due to sampling limitations. In that case one option is just to plot everything (including AERONET) as a seasonal rather than monthly time series. This would also remove the issue of different expected extreme which the authors mention in captions.

Re: Seasonal comparison rather than monthly comparison in time series is definitely more concise for reading, but that will lose some details. Hence, we just keep monthly time series for each except ATSR. On this basis, if we put everything into one plot, this plot will be too messy to read like following figure.



The focus of this paper is to investigate consistency of ATSR and AVHRR, while AERONET and MODIS are chosen as reference data. Hence, we make comparison of MODIS and AERONET separately to see the performance of MODIS. When it is confirmed, time series comparison will be made with AERONET or MODIS as reference separately. Please refer to fig. 4, 5, 9, and 10 in new version of manuscript. Those time series are looking much better than before.

Figures 7, 8: these are interesting, but I feel they should be expanded to make the paper more complete. They show that the monthly solar broadband-based AOD is a reasonable proxy for AOD from AERONET or MODIS. But why not also include scatter plots like figure 7 for AVHRR and AATSR? And why not include the full time series, as well as ATSR data, in figure 8? Basically the paper is about a 1983/7 and onward AOD data set over Beijing, but nowhere is a full time series of these data actually shown in the paper.

Re: As last comments, we didn't put everything into one full time series plot. Instead we made two time series, one is to validate BEM AOD and the other is to be as reference to validate AVHRR and ATSR. Please refer to fig. 6 and 8 in new version of manuscript.

Conclusions: The paper also builds on other recent work by the de Leeuw group which assessed a combination of the ATSR and MODIS sensors to get a longer-term AOD over China. This is mentioned a few times early on, but I think the conclusion should be expanded to put these studies in perspective with each other, and give some more specific thoughts/suggestions on how best to combine data records. Overall this paper has some good points but suffers from not always following through to give the bigger picture (i.e. when you have finished your investigation of the long-term climate data record, as promised in the title, what does it actually look like) ?

Re: Together with other two reviewers' comments, we added more conclusions.