Atmos. Meas. Tech. Discuss.,
doi:10.5194/amt-2020-336-RC1, 2020

Atmospheric Measurement Techniques Discussions

Interactive comment on "Integration of GOCI and AHI Yonsei Aerosol Optical Depth Products During the 2016 KORUS-AQ and 2018 EMeRGe Campaigns" *by* Hyunkwang Lim et al.

Anonymous Referee #1

Received and published: 9 November 2020

This paper merges and analyzes aerosol optical depth (AOD) data from four data sets (two sensors – AHI and GOCI – each with two different algorithm versions) by two methods (simple mean and maximum likelihood) during two field campaigns in East Asia. Individual and merged data sets are evaluated against Sun photometer observations (more dense than usual due to the field campaigns); statistics of the individual product comparison are also used to inform the merging process for maximum likelihood.

The paper is relevant to the journal and the special issue. The topic is important: we have a lot of satellite AOD data sets now and the question of merging comes up increasingly often. It is also nice to see the geostationary data here; this is a novel aspect

and these new sensors offer temporal coverage unavailable from polar platforms (as the authors point out). So this is all good. The quality of language is ok: the authors have done a good job considering their native languages are not English, but some copy-editing will be required. This can probably be handled by the journal. As a result I have only made language comments when it relates to technical issues.

Some of the analysis is unclear, in particular, relating to the bias correction step (see later comments). I also found the organization of the paper hard to follow: a lot of different merging results were presented but the main message is not clear and I am not sure how well these results could be generalized to other time periods (outside of these field campaigns) or other data sets. Right now it is hard to tell if this is more a paper about these field campaigns, or these retrieval algorithms, or merging in general, because it's not focused/in depth enough. As a result I recommend major revisions to address these issues. My main recommendations relate to streamlining the analysis and discussion, and using more modern merge techniques. I would like to review the revised version. Specific comments in support of my recommendation are below:

1. Line 31: "affect radiative energy" should probably say "affect Earth's radiative energy balance" or "affect solar and thermal radiation" as the current wording feels a little odd.

2. Lines 38-49: there are long citation lists here, with some repetition, and not really much discussion. I suggest consolidating this. We know there are many AOD retrieval algorithms, there's no value in listing a bunch of references unless they are discussed in more detail (as in the examples in the next paragraph). This is an issue elsewhere in the introduction as well, but especially here.

3. Line 50: DT is not one algorithm. It is two algorithms: one for land, one for water. They have the same name, but the assumptions (e.g. aerosol properties, surface reflectance) have nothing in common and even the channels used are different. This should be corrected.

4. Lines 111-125: here the authors describe a number of approaches which have been

used to merge AOD products. Given the sophistication of many of these methods, why are such simple methods (i.e. simple mean, and MLE – which is essentially an uncertainty-weighted mean) used in the present study? Why not use something more state-of-the-art? This paper seems a bit of a missed opportunity to study whether more advanced data fusion approaches as cited in these lines do any better than simple mean or MLE. The authors might consider trying to add a more advanced technique.

5. Section 2: I did not find a clear description of what wavelength(s) AOD is reported at in this analysis. From a few figure captions I think 550 nm, but this seems to be the only mention in the text. This should be stated clearly for each data set used, along with any method for spectral interpolation applied.

6. Line 161: is 0.02 mg/m3 correct? This seems unrealistically low. I was surprised so looked through the Yamada paper cited and did not find this number supported. It looks (e.g. their Figure 2) that most of the time, for their limited domain, the climatological value is 0.1-1 mg/m3. However there is considerable variation. So using 0.02 seems wrong, and having no spatial/monthly variation also seems like it would introduce seasonal biases.

7. Lines 234-244 and Table 1: This is where things get messy for me. I feel there are too many comparisons (7 merge tests, 4 un-merged data sets) and it gets difficult to remember which combinations of algorithm acronym belong to each data merge acronym without going back and forth to the table each time. Further, I am not sure that the split as presented enables the analysis authors want to do. It is complicated because we are splitting between not only different merge types, but also different numbers of sensors (as GOCI has a smaller disk), and also different observation regions (and we know aerosol and surface characteristics, as well as retrieval errors, are probably different in these regions). It is not comparing apples to apples. After reading the paper several times, I'm still not very sure what the message is and how general this recommendation might be. I wonder if it makes more sense to drop some of these experiments and focus only on the ones involving the GOCI disk in order to have a clearer picture for the

analysis (consistent spatial domain, smaller number of comparisons, smaller region to map to make figures easier to read). Maybe doing this, and adding a more advanced merge method (see earlier comment), would give an analysis which is easier to follow and of broader interest. Having all the 11-panel figures which look mostly quite similar is hard to follow.

8. Section 3.4: this section doesn't seem to actually explain how the bias correction was done. More detail is needed. Also, I don't see the evidence that retrieval errors do follow a Gaussian distribution: there is no Gaussian distribution comparison shown in Figure 1. This could be demonstrated better by e.g. a QQ plot. Further, it could be that there are multiple populations in here and it looks reasonably Gaussian on average, but not for subsets of the data.

9. Line 264: Sayer (2013) does not show AOD follows a lognormal distribution. Perhaps the authors are thinking of Sayer and Knobelspiesse (2019)? https://acp.copernicus.org/articles/19/15023/2019/

10. Sections 4, 5: these mostly just describe the figures and again, because there's a lot of maps and scatter plots which look very similar, it is hard to pick out the main message. This supports my idea to pick which experiments and parts of the data are most important and focus on those. In my view the figures should support the text; the text should clearly offer explanations and recommendations and not just describe the figures. I don't have many more specific comments on these for that reason.

11. Tables 2, 3: these are a bit of a sea of numbers. It is hard for the reader to parse them and extract the main message. If the variation between entries is important, perhaps these should be figures instead. Also, "NaN" does not belong in a table like this. If there were no data, leave it blank or put a "-". NaN is computer code.

12. Figures 5, 7: I recommend the regression fits be removed here. As the authors note, AOD is close to lognormal. Also, the AOD error is dependent on AOD. Also, the fact that there are NDVI dependences of retrieval errors means that there are multiple

populations of data with distinct characteristics here. All this means that the regression used is statistically inappropriate. It should be removed in order for the paper to be correct. I do not believe the regressions are vital to the discussion anyway.

13. Figure 9: can uncertainty bars be added here? It is hard to see whether these differences are real or within sampling error. Also, the x axis should be checked. While NDVI below zero is mathematically possible, it is not realistic except for water bodies or cloud-contaminated pixels. I am surprised that values seem to vary between -0.3 and +0.4 or so. Even deserts have an NDVI around 0.1-0.2, and vegetated areas often above 0.5. I wonder if there is perhaps a bug, a definition difference, or a serious spectral error in the surface reflectance model producing these values. This should be checked.

14. Figure 10: this has vertical bars but they are not explained. Is this standard deviation, standard error, or something else?

15. Conceptually, I also have an issue with using AERONET to train a bias correction and then evaluating the bias-corrected data against AERONET. Of course this will look better than the original products. I am not sure of the best way around this though. Again, streamlining the number of comparisons made in the paper will make it more readable and allow a better understanding of the advantages of the methods.

Interactive comment on Atmos. Meas. Tech. Discuss., doi:10.5194/amt-2020-336, 2020.