Atmos. Meas. Tech. Discuss., 5, C1792–C1806, 2012
www.atmos-meas-tech-discuss.net/5/C1792/2012/

Atmospheric Measurement Techniques Discussions

Interactive comment on "Accuracy assessment of Aqua-MODIS aerosol optical depth over coastal regions: importance of quality flag and sea surface wind speed" *by* J. C. Anderson et al.

Anonymous Referee #1

Received and published: 17 August 2012

General comments and recommendation:

This paper presents a validation of MODIS aerosol optical depth (AOD) against AERONET data, with a focus on coastal regions. Various statistical techniques are used and show that MODIS compares less well against AERONET for coastal sites than non-coastal sites. Then, dependence of results on retrieval quality flag, and the effect of introducing a correction term based on wind speed to the MODIS retrievals, are discussed.

I appreciate what the authors are trying to do here. However, I feel that the study falls short of their goal. Many of the results from the analysis are not new: there is a lot of retreading of old ground, except with a focus on coastal areas rather than global results, and in most cases it turns out the "coastal only" results are similar to the "all sites" results reported in previous papers. I am not sure what the end goal of the authors is beyond this work. If it is to create a bias-corrected MODIS product for trend analysis/data assimilation, have they considered just using that which already exists and is documented in multiple studies they cite? In several cases the statistical techniques used are not appropriate. Some statements made in the paper are incorrect and could be misleading to a reader unfamiliar with some aspects of MODIS data. Further, I think the authors focus on statistical significance (which is of course important) almost to the exclusion in some cases of scientific significance, and miss the point a little with regard to some aspects of that (e.g. they show some differences which are statistically significant according to their analysis but probably negligible for many applications). These points are expanded on in my specific comments, below.

One further aspect, which some (maybe all) of the authors will be aware of, is that MODIS Collection 5 data will be superseded in the coming months by Collection 6. One of the main conclusions of their article, supported by prior work that they cite and has been known for several years now, is wind-speed dependence of error in the MODIS AOD over ocean. This is one of the factors which has been addressed and should no longer be an issue in the Collection 6 product. In this sense the authors' validation exercise and suggestion for improvement has arrived several years later than ideal, since it is soon going to become outdated. Other factors such as calibration and cloud-screening improvements will also mean that the other conclusions are likely to change quantitatively. The Collection 6 updates and impact have been presented at numerous national and international meetings for more than a year now, and the

authors acknowledge R. C. Levy so have been in contact with the MODIS science team, and surely must have been aware of this issue. I appreciate that a dataset being updated is not the fault of the authors, but still it lessens the scientific value of the study, and raises the question of whether it is worth publishing the results or waiting a few months and rerunning the code with the Collection 6 product–which would give a much more useful study.

For these reasons, I recommend the manuscript undergoes major revisions and then a second round of peer-review. In brief, I recommend three main strands for revision, justified in the specific comments. Firstly, due to limitations of the current MAPSS software, it would be much better if the analysis were done with the "central point" rather than "box average" method. Secondly, the authors could request MODIS Collection 6 data from the MODIS science team (or wait for general release), and repeat the analysis with that, which would be worthwhile. This will cause a delay but the quality of science should be most important factor in these decisions. Additionally, using Collection 6 would doubtless increase the readership of the paper. Thirdly, a lot of the repeated work from previous studies (quality flags, wind speed) could be omitted or shortened, and then some new section showing a scientific application of the filtered/bias-corrected data could be added. The paper as it stands now is not too long, and that would add something more original to the study.

To my best knowledge the lead author is a graduate student and this is their first publication submitted for peer-review in a journal. I do not wish to be discouraging, as I think this type of study is an important one, but would much rather see a heavily-revised paper with strong and rigorous statistical treatments and discussion, than have a weaker paper published. There is potential for a very nice study in here. I am sure that being aware of the types of issue raised in this review will benefit the student and their studies in the long run. Please do not feel disheartened–I know this

is quite a lengthy review but I think you can get a good paper out of this, with some more work and correction of errors and misconceptions, and I want to see it.

Specific comments:

Abstract: This is quite long and some of the text is not needed for an abstract. The text may change based on the revision of this paper, but my suggestion based on the current version would be:

"Coastal regions around the globe are a major source for anthropogenic aerosols in the atmosphere, but the underlying surface characteristics are not favorable for the Moderate Resolution Imaging Spectroradiometer (MODIS) algorithms designed for retrieval of aerosols over dark land or open-ocean surfaces. Using data collected from 62 coastal stations worldwide from the Aerosol Robotic Network (AERONET) from 2002–2010, uncertainty assessments are made for coastal aerosol optical depth (AOD) retrieved from MODIS aboard the Aqua satellite, from the Collection 5 dataset. It is found that coastal AODs (at 550 nm) characterized respectively by the 'Dark Target' land algorithm, ocean algorithm, and AERONET all exhibit a log-normal distribution. After filtering by quality flags, the MODIS AOD is highly correlated with AERONET (with R2=0.8), but only fall within the expected error envelope greater than 66 % of the time for the land algorithm. Furthermore, the MODIS AODs show statistically significant discrepancies from their respective counterparts from AERONET in terms of mean, probability density function, and cumulative density function. Without filtering with quality flag, the MODIS land and ocean AOD dataset can be degraded by 30–50 % in terms of mean bias. Overall, the MODIS ocean algorithm overestimates the AERONET coastal AOD by 0.021 for AOD < 0.25 and underestimates it by 0.029 for AOD > 0.25. This dichotomy is shown to be related to the ocean surface wind speed and cloud contamination effects on the satellite aerosol retrieval. The Modern

Era Retrospective-Analysis for Research and Applications (MERRA) reveals that wind speeds over the global coastal region (with a mean and median value of 2.94ms-1 and 2.66ms-1, respectively) are often slower than 6ms-1 assumed in the MODIS Ocean algorithm. An empirical scheme for correcting the bias of AOD retrieved from the MODIS Ocean algorithm is formulated and is shown to be effective over the majority of the coastal AERONET stations."

Page 5207, lines 17-26: You may consider the paper by Hsu et al (ACPD, 2012), which looks at AOD trends from SeaWiFS both over land (as you note, previous studies were ocean only) and ocean: Hsu, N. C., Gautam, R., Sayer, A. M., Bettenhausen, C., Li, C., Jeong, M. J., Tsay, S.-C., and Holben, B. N.: Global and regional trends of aerosol optical depth over land and ocean using SeaWiFS measurements from 1997 to 2010, Atmos. Chem. Phys. Discuss., 12, 8465-8501, doi:10.5194/acpd-12-8465-2012, 2012.

Page 5208, line 10: I suggest "uncertainty" rather than "accuracy" here, as this is the uncertainty confidence envelope, while "accuracy" (and "precision") have specific technical meanings which are different. This should be checked throughout the manuscript.

Page 5208, lines 12-14: This discussion in the context of the MODIS uncertainty is not directly relevant and misleading. The uncertainty envelope is defined only for an instantaneous MODIS retrieval, and cannot be quoted in the context of trend analysis. Unless you know the systematic and random components of the uncertainty, and how they change over time, you cannot propagate the instantaneous uncertainty into the trend. And these things are not known quantities. For example, if there were a bias of 0.1 in MODIS AOD all the time but no other source of error you could still use it for trend analysis because the uncertainty has no time dependence, even though the total uncertainty might be larger than the trend. Additionally, for trend analysis, there are a whole other set of factors, such as temporal compositing strategy, spatial/temporal

sampling, etc, which play a role and determine how easily any trend can be identified. This sentence must be reworded.

Page 5208, line 29: the official name is "Dark Target" rather than "Dark Land"; it would be preferable for the authors to use this, to minimize confusion, but not a critical issue so long as they are self-consistent.

Page 5209, lines 22-25: This is another example of an incorrect statement which is misleading and must be corrected. There is no loss of quality information in the combined dataset, precisely because it is a union of the land and ocean datasets, the same quality flags apply as for the separate datasets!

Page 5210: The discussion on wind speed is a bit long, particularly as, as you note, it has already been analysed and known about in multiple studies for several years now. If you wish to include this analysis in your paper, it would be sufficient just to cite those studies and say you will examine whether the same results hold over coastal areas as over the ocean as a whole. As mentioned previously, doing the analysis with Collection 6 data and seeing whether the wind-speed dependence has been removed successfully would be more useful.

Section 2.3: The MODIS ocean product provides two AOD datasets: from the best-fitting aerosol model, and from the average of several well-fitting aerosol models. Which is used here? This should be stated.

Page 5211, line 2: I suggest Levy et al (ACP, 2010) here as the more useful reference as it shows the drift in Terra validation statistics over time.

Page 5211, line 21: Missing space in ‘vegetatedsurfaces’.

Page 5211, line 22: Missing space in ‘wavelengthover’.

Page 5211, line 24 and page 5212, top: The acronym ‘LUTs’ should be defined as first use. I think this paragraph can probably be deleted and the algorithm papers just referenced. You are only describing the land algorithm here, not the ocean algorithm (although you don’t state this in the text), and the information you give here is not used later in your discussion.

Page 5212, line 13: ‘quality assured’ is not meaningful here; what do you mean? I suggest deleting these words.

Page 5212, line 21: AEROENT should be AERONET.

Page 5212, lines 21-24: This sentence is hard to follow and be reworded. Just say that using quality flag 3 over land and 1, 2, or 3 over ocean should give agreement within EE 66 % of the time on a global basis. The way it is written makes it sound as if each retrieval is validated and then assigned a quality flag, which is obviously not the case, and could be misleading.

Page 5213, line 1: You should reference the AERONET cloud screening paper here: Smirnov, A., Holben, B. N., Eck, T. F., Dubovik, O., and Slutsker, I.: Cloud-screening and quality control algorithms for the AERONET database, Remote Sens. Environ., 73, 337–349, 2000.

Page 5214, lines 1-7: Are you just using all AERONET sites? There are some (e.g. Mauna Loa, Izana) which you should exclude from your analysis, as they are e.g. high-altitude sites which are not representative of their larger region on MODIS retrieval spatial scales, so not a useful validation site for MODIS data. Using everything blindly will make it more difficult to draw meaningful conclusions as discrepancies will arise for reasons other than retrieval error. Some years ago Stefan Kinne (MPI-Hamburg) was compiling a list of AERONET sites he thought were representative of their larger-scale environment. I don’t know whether that list was published as part of any study but it could be worth asking him. Certainly you should consider only a subset of the available AERONET sites.

Page 5214, lines 15-19: you say there is “little difference” between the two MAPSS analysis methods yet elsewhere you state as “significant” differences which are of order 0.01-0.02, i.e. within or near the AERONET uncertainty. You should quantify here exactly how different the two methods are.

Also, as you note, the MAPSS “average” method does not unambiguously look at the effect of quality flags because it takes the mode (rather than subsetting for each quality flag). This appears to be a limitation of MAPSS. If you are really interested in looking at the effect of quality flags, as you do later, this implies you should really use the “central” rather than “average” method. I am sure data volume would be sufficient.

Page 5215, lines 8-11: I mention this issue here but it applies at several later points too. I know ordinary least squares (OLS) linear regression is popular in our community. But it is really the wrong thing to do here, and your fits will be skewed. This is a well-known issue presented in statistical textbooks. Being a popular technique does not mean it is a good technique to use. The assumptions for the regression you have done are that the relationship is linear (maybe ok, although Figure 3 shows nonlinearities) and

that the noise about the linear relationship is Gaussian and the same size across the range of the independent variable (here, AERONET AOD). As you note several times already in the manuscript, the MODIS uncertainty has a dependence on AOD, so you have already stated in your manuscript one reason why OLS regression is not an appropriate technique (i.e. scatter at low AOD and at high AOD are different)! In addition, the uncertainties at low AOD are more likely to be biased high than low, because negative AOD retrieval is not permitted over ocean and you say later you throw negative points out over land. This will result in the linear fits being biased to overestimate at low AOD and underestimate at high AOD–exactly as observed in parts of this study and others. Sampling is also extremely non-uniform along the AOD axis, because AOD distributions vary (as you note) approximately lognormally. You should not use OLS regression but instead another technique (there are several options) which is statistically appropriate. Just because others have published with such erroneous techniques does not mean you should fall into the same trap. In fact doing it properly would hopefully serve as an inspiration to others in the future.

Page 5215, lines 11-15: It would be interesting to see the distributions of the bias. The mean is a handy statistic but if you have outliers, skewedness and so on, other quantities such as median and standard deviation/interquartile range can be more informative. I suggest showing these distributions in the revised manuscript. The distribution of error, and/or the distribution of error relative to the expected error (which one would hope is Gaussian but could well not be) would be good to see.

Page 5216, lines 6-8: This is not a good thing to do. If you are throwing away negative retrievals, which you know are low-biased, you will therefore skew your analysis towards reporting a more positive bias than is really the case. I understand that you can't do a lognormal fit with negative values, so, ok, you can throw out those 400 points for this, but they should be included for the rest of the analysis of the paper.

Also, what proportion of your total sample is 400 points? Sampling is given in Table 2, but mention the relevant number in the text here too.

Page 5218, lines 7-23: See comments about OLS previously; an alternative method should be used. These R2 and regression fits would be much clearer, and the paragraph more readable, if presented in a table.

Page 5219, lines 11-12: This sentence is jarring. You are basically saying you use 0.25 as a threshold because another paper used 0.2. I don't think you need to justify a choice of 0.25 as something in that region fits with common sense. So, I'd either delete the mention of Levy et al (2010) or else adopt their 0.2 threshold (which I expect won't change things much).

Pages 5519-5521: here you commence looking at differences and calculating statistical significance. But remember that AERONET AOD has an uncertainty of 0.01-0.02, which is of a similar size to the mean differences you are claiming as "significant". In this light, can you say it is scientifically significant? On page 5220-5221 you say "MODIS is not very accurate in modeling the actual nature as represented by AERONET" and that "MODIS does not model the actual nature represented by the AEROENT AOD observations". Aside from the previous point of whether "accurate" is the correct word here, this conclusion is not necessarily warranted from a scientific point of view. The relevant question is whether the differences are important for a given application (as mentioned in the General Comments, I am not sure what the end goal of the authors is from this study). Are we interested in typical values or extremes? It might be that one is consistent between the two datasets and the other not. The CDFs in Figure 6 look very similar, with much of the difference contributed by points off the left-hand side of the plot (also see later comment on that figure)–and as you say in the text you are throwing some negative values out over land. Would

just removing the average wind-bias offset from the MODIS ocean AOD make the two statistically consistent? It looks like it might help, certainly. I suspect for many applications these differences are not important. Be careful not to confuse statistical and scientific significance. You can detect two things are statistically different (e.g. the distribution of sizes of grapes in two bags) but it does not mean it is important (both may provide a tasty snack). The discussion should be extended and these aspects discussed.

Remember also that you are not looking at the MODIS or AERONET data directly here. You have a spatial average of MODIS and a temporal average of AERONET, and averaging is going to change the shape of the PDFs/CDFs, dependent both on the noise in each dataset and also on the spatial/temporal variability of each dataset. You don't really mention this aspect, although it is discussed in several other satellite AOD validation studies. So this analysis and conclusion is misleading. This is another reason it would be much better to use the "central" MAPSS method: you can look at quality flags directly, and it removes the effect of spatial/temporal averaging from the comparison of PDFs/CDFs. You claim that in your analysis you are going beyond previous studies. As I said in the general comments, there is a nice study buried in here, but to get to it you are going to need to redo the calculations and add to the discussion, particularly in this section. Otherwise it is not really adding anything useful over previous studies.

Section 4 (page 5221, lines 19-24): This is the same misleading statement the authors make earlier and must be corrected. Again, the LandAndOcean quality flag is the same as that for the Land dataset for pixels over land, and the same as that for the Ocean dataset for pixels over ocean. I question the value of this section and think it can be removed or summarized in a single sentence, unless you repeat the analysis with Collection 6 data, or do something else which is really new. The MODIS team

recommends applying the quality flags, precisely because if you don't the data are of lower quality. This has been shown in previous studies, and doesn't need to be rehashed here. Also, the point about OLS regression applies again here.

Page 5223: Yes, Zhang and Reid (2006) was about the Collection 4 product. However, Shi et al (2011), which you cite elsewhere, is a similar analysis for the Collection 5 product (reexamining clouds, wind etc). So, this has already been extensively investigated on a global basis (and with similar results to the coastal analysis) to create their group's data assimilation MODIS products, which include this additional cloud filtering. This section is retreading old ground. This is another aspect which would be much more useful with Collection 6 data, for which the analysis has not yet been done, and cloud contamination issues should hopefully be smaller.

Page 5224-5226: This is retreading old ground from the Zhang/Reid group papers, and does not provide new insights. I suggest removing it, or waiting and repeating the analysis with Collection 6 data.

Page 5227, lines 1-8: If 46 out of 62 coastal sites have a significant relationship, the more interesting question is: what about the 16 which don't? Which are they, and how/why are they different? Perhaps the dominant aerosol source is not marine at those 16? I suggest discussing these sites in more detail, which may provide new insights concerning these regions which could be useful for other studies. They are shown in Figure 7 but there is no interpretation offered, and there should be.

Page 5227, lines 11-13: You mention "complex surface characteristics". This is true for e.g. the 1 km pixels which straddle the land/ocean boundary, or even the 10 km retrievals which include water, beach, vegetation, urban areas, etc. But since you are

averaging a 55 km box, much of these areas will include the same types of terrain that you see in non-coastal areas. So I don't think the statement you make in this sentence is a fair assessment of what is tested in this work. Again, using the "central" MAPSS method rather than "average" would be an improvement because you specifically would be looking at these complicated heterogeneous regions.

Pages 5227-5228: The rest of the conclusion is somewhat brief. In the end you state that bias correction improves the agreement between MODIS and AERONET and should be done for trend analysis and data assimilation. Well, this is what the Zhang/Reid et al analyses have shown for some years, which you cite, and has been used to produce their data assimilation MODIS dataset. So, what is new from this study? Again, using Collection 6 data, and including some further analysis where you make use of your bias-corrected dataset, would make this much more interesting and useful.

Tables 1, 2: Mean bias, as mentioned earlier in the review, is not a very useful metric by itself. What about the median bias, standard deviation of bias, etc? Such other quantities would also be worth putting in the tables.

Figure 1: I think there is an inconsistency between figure labels and text here (e.g. B is labeled LandAndOcean but the fourth, not second, plot listed in the caption). I would suggest it would be more informative to include the numbers from this plot (plus median AOD, as well as mean) in a table, and instead plot these relative frequency distributions on top of each other with lines. That would enable a more direct comparison of exactly where in AOD space the distributions are different. Also, these bins are somewhat coarse: I suggest narrowing them to see whether there is any finer-scale structure.

Figure 2: How do these results change if you use finer bins? Since you do not have a large number of bins, it is easier to find a statistical agreement between observed and theoretical lognormal distributions. It would be more convincing if the bin size was smaller and you still found statistical significance.

Figure 3: These would be clearer if you plotted scatter density plots, e.g. Figure 1 of the Levy et al (2010) MODIS validation paper. That lets you see where the bulk of the data actually are, rather than a cloud of points. Again, the OLS linear regression should be replaced by a more appropriate technique. Look at any of the top panels, especially the top-right. It is obvious that the OLS linear fit is not a good model for what is happening at high AOD.

Figure 4: This is a nice way to show things.

Figures 6, 9D: I have some concerns with these. First, if I understand correctly, these show the lognormal fit distributions of the data, rather than the actual data themselves. Given the departures from the lognormal fits (very evident in e.g. Fig 2D), I think you are introducing a non-negligible uncertainty by using the approximate fit distributions, especially given your criterion for what is a "significant" difference is very small. So, this is somewhat misleading and could be affecting your conclusions. It would be much better to use the actual distributions. I understand these are not continuous distributions, but you could plot the CDF of the binned data here, which would be good enough (particularly if you take my suggestion of narrower bins). That would be a much fairer treatment. Or, you could do this type of analysis using QQ plots instead of the CDF. As it stands, I worry that comparing the CDF of distributions which your data only approximately match, as opposed to CDFs of your actual data, is just making things difficult.

Secondly, the inference from these figures is that about 15 % of the time the AOD is less than 0.01 in all datasets. This is surprising. I would check your code. Your Figure 2 does not support this (almost nothing below 0.018), and I also looked at AERONET data for a few sites (Lanai, COVE) and found almost no points with AOD at 550 nm < 0.01. I note your Figure 2 is natural logarithm, while Figures 6 and 9D appear to be base-10 logarithm. Perhaps somewhere in your calculations one was incorrectly used, introducing the error.

Figure 7: There are several island sites which appear to be missing from this plot, e.g. from cursory examination Arica, Lanai, Midway Island, Ascension Island, Reunion, Tahiti, plus the Australian coastal sites (Darwin appears in the top two panels, but not the bottom two; this should be checked). Perhaps more. Is this just because there were no valid land retrievals for them? These are fairly well-established AERONET sites with lots of data, and some are on islands significantly larger than the 10x10 km nominal MODIS retrieval size. I suggest checking up on this.

Interactive comment on Atmos. Meas. Tech. Discuss., 5, 5205, 2012.