

Interactive comment on “MISR Research Aerosol Algorithm: refinements for dark water retrievals” by J. A. Limbacher and R. A. Kahn

A. M. Sayer (Referee)

andrew.sayer@nasa.gov

Received and published: 13 August 2014

I am posting this review under my name to disclose that I collaborate in research with the authors of this study. However, I feel that I am able to provide an objective review of their manuscript.

The authors describe refinements to an algorithm for remote sensing of aerosol properties over water from Multiangle Imaging Spectroradiometer (MISR) measurements, with a focus on spectral aerosol optical depth (AOD) and Ångström exponent (ANG). The algorithm is not the MISR Standard Algorithm (SA) data product, which is freely downloadable, but the authors' own Research Algorithm (RA), which is a more flexible

C2153

and improved version of the SA that has been used for case studies of various aerosol events. The refinements made to the RA in this study are generally physically-based and justifiable, and result in an improvement of statistics (compared to the SA and previous version of the RA) when spectral AOD and ANG are compared with Aerosol Robotic Network (AERONET) island and Maritime Aerosol Network (MAN) ship-based validation data.

General comments and recommendation

The topic is appropriate for AMT. It is of interest for two main reasons: first as an update and reference for the RA for future applications, and secondly as a reference for (presumably) eventual incorporation of the algorithmic improvements into the MISR SA, where they would benefit the scientific community more directly by being available for broader use.

The paper is largely scientifically sound, although I have some comments and suggestions (provided below on a section-by-section basis). Most of these are requests for clarifications in the text. MISR has historically had unique terminology associated with it and I think that this could in general be stated more clearly. Additionally, I have a few suggestions for alternative/additional analyses: specifically, changing the way stratification by chlorophyll content is done, more discussion of 'best' vs. 'averaged' retrieved aerosol mixtures in the AERONET comparison, additional processing to illustrate larger-scale effects and future SA application, and discussion of retrieved quantities besides AOD/ANG. I therefore recommend revisions (which are mostly minor aside from some larger-scale processing); I feel the paper can be accepted for publication in AMT if these suggested revisions are made. I am happy to review the

C2154

revised manuscript if necessary.

Specific comments

Abstract

P7838 line 3: The MISR acronym should be defined on first use here.

P7838 line 4: I think the authors should just say 'uncertainty' or 'error', as 'bias' is a component of the total uncertainty/error.

P7838 line 12: Presumably spectral 'calibration' uncertainty.

P7838 line 22 (and throughout the manuscript): The authors describe error bounds with terms like 'within 0.05 or 20%'. What I think they mean is 'within the greater of 0.05 or 20%', as used in some of their other publications. This distinction is important, to decrease the chance of confusion, so I suggest this be changed throughout.

P7838: Terms like 'blue', 'green', 'midvisible' etc should be avoided here and explicit wavelengths given. I think it is fine to refer to the colour names for the wavelengths in the main text, as the name/wavelength mapping is given in the Introduction, but in the Abstract it's better to be explicit.

C2155

Introduction

P7839 lines 2-14: The MISR launch date and platform, and swath width, should be provided with the other sensor information in the Introduction.

P7839 lines 15-25: There is a lot of MISR-specific terminology (e.g. 'region') as well as terms which mean different things in different contexts (e.g. 'pixel'), here and elsewhere in the manuscript (e.g. 'subregion', 'mixtures'), which would benefit from being defined explicitly early on in the Introduction, as many readers will be less familiar with MISR than other datasets.

P7839 line 26: The authors draw the distinction between the RA and SA here. But I think it would be helpful to add an explicit statement somewhere in the Introduction (and again in the Conclusion) to state that RA results have not been used to process the whole MISR record, and are not available to download from the Langley ASDC. People may gloss over this or misread, so it is better to over-emphasise this fact to help minimise the chances of people getting the wrong impression. If you search the literature you often see incorrect/outdated citations to datasets so I think it's best to nip this in the bud.

P7840, lines 1-4: As this paper is exclusively about dark-water retrievals, I don't think it is necessary to mention the RA over-land surface BRDF treatment here, and would delete this sentence.

P7840 lines 5-6: The algorithm is referred to as a 'dark water' algorithm, although it isn't actually stated what this means. For example, is this determined dynamically during processing, is it user-defined, or is there a geographical constraint used? This

C2156

should be stated. It would be good to show e.g. a map showing where 'dark' and 'non-dark' water occur in the MISR SA/RA, so we can see the geographic extent of the algorithm being discussed (this will be particularly relevant if/when these improvements are ported to the SA). For example, later on the authors state that only AERONET island sites are used due to other issues at coastal sites. But are these coastal sites counted as 'dark water' normally or not? What about large inland lakes or seas? What about coastal but deep water?

Section 2

P7841, line 17: I don't think this parameter A_0 is mentioned after this point, so the variable letter can be deleted for clarity.

P7841, line 19: The AERONET acronym should be defined on first use.

P7841, line 23: I know what the authors mean here, but it would be cleaner to say 'aerosol mixtures' or 'aerosol optical models' rather than just 'mixtures'.

P7841, line 27: Why is cloud contamination not listed as the fourth factor? I suppose I am asking why say there are 'three main factors ... along with possible cloud contamination' and not 'four main factors'? I suggest changing this. I would also suggest another (fifth) factor is structural limitations of the radiative transfer code and method of finding the retrieval solution. For example, I understand the SMART code is scalar rather than vector. While this is perhaps less of an issue for the bands used in the dark water RA it will still make some difference (especially as the authors recommend calibration adjustments as small as 0.75% later in the paper), and there are other

C2157

radiative transfer approximations (e.g. plane-parallel, 1D radiative transfer, aerosol mixing, homogeneous scene) and algorithmic methodologies (e.g. method of finding the solution) which will introduce some error (that may not necessarily be unbiased). So I think it is important (here or elsewhere) to mention these more 'structural' aspects of the radiative transfer/retrieval algorithm as contributing to retrieval uncertainty.

P7842, lines 9-11: Is there a reference for the scatterometer data? What is the uncertainty of the wind speed, and how do its resolution/spatial completeness compare with MISR? As the authors know, wind speed uncertainty can become one of the main sources of uncertainty for AOD retrieval in low-AOD open-ocean areas.

P7843, lines 8-9: I think what is being said here is that the method of finding the retrieval solution is not an iterative minimisation of a cost function (as many other algorithms use), but a brute-force method on the fine-resolution lookup table. I think it's worth stating this more directly, or if I have got the wrong impression, rephrasing.

P7843, line 12: I think the denominator of the upper sum should be typeset as ρ_{err}^2 rather than ρ_{err}^2 , because the subscript 'err' is part of the variable name. Note this is different from the case of χ^2 because χ^2 is itself a widely-used entity distinct from its attached subscripts. I could be wrong on this though.

P7843, line 14: What values are used for the band and camera weights w ? I don't think this is mentioned in the paper. If they are weighted equally, this term can be removed from Equation 1 and the discussion for simplicity. Also, is there some advantage to doing it this way rather than simply change the assumed uncertainty ρ_{err} (which is effectively a weight on each measurement)?

C2158

P7843, lines 21-25: These additional metrics should be explained more clearly and defined explicitly. I infer χ_{geom}^2 and χ_{spec}^2 are calculated using Equation 1, but omitting loops over wavelength and camera respectively. Is that correct? If so, which reference wavelength/camera are used? Or is the calculation something else? It would probably be clearest just to provide equations. Also, is χ_{maxdev}^2 simply the largest individual (band/camera) contribution to χ_{abs}^2 ? Weighted or unweighted? From P7844 line 9 I think this is the case, but it would be better to have the definition of this variable in one place rather than split across two paragraphs.

P7843, line 26 and P7844, lines 1-2: The blue and green MISR bands are not used in the algorithm because of stated larger uncertainties in ocean surface reflectance modelling. However, no justification in support of this statement is provided. Given that in Section 3.2 the authors improve the ocean surface model, this would have seemed a good opportunity to see whether the uncertainties are still large enough to preclude the use of these bands (and it would provide some quantitative evidence either way). If they're not, then it might provide some additional skill in aerosol mixture selection (given the typically increased contribution of aerosols to the top of atmosphere signal at longer wavelengths). However, as far as I can see, this is not discussed. I think it should at least be mentioned (probably somewhere in Section 3.2 later on). Surely including the channels, even if with low weights in the χ^2 tests (to account for the larger uncertainties), is better than not including their information at all? And how do the authors know that the largest error sources at these wavelengths are related to the ocean surface reflectance and not, say, using a scalar radiative transfer code, or aerosol optical model/vertical distribution assumptions?

P7844, line 6: Is this what is used to define 'dark water'? See my earlier comment on this subject.

C2159

P7844, line 18: Is the word 'absolute' needed here? When dealing with real numbers an absolute value cannot be below zero.

P7844, line 19: Do the authors mean 'brighter than expected' rather than 'very bright'?

P7845, line 9: I assume the authors are using AERONET level 2 version 2 direct-Sun data, although this is not stated. It should be, and appropriate references provided. Holben et al. (1998), cited elsewhere in the paper, is good for AERONET as a whole, but Smirnov et al. (RSE, 2000) describes the automatic cloud flagging algorithm for direct-Sun data, and Holben et al. (SPIE, 2006) describes quality control information for version 2, which may be of relevance.

P7845, lines 9-11: This is the type of information which I think should be provided in the introduction to MISR data, when terms like 'region' are first used, rather than here.

P7845, lines 11-12: Why do the authors think that AOD measured by MAN has greater temporal variability than that for AERONET island sites? Is there evidence for this? I would have thought the opposite: if the MAN cases are open-ocean then they might be more spatially/temporally homogeneous on these scales than the islands (as islands may have some additional local sources).

P7845, lines 17-24: These sentences deal with AERONET site selection. Coastal sites are excluded, although it is not clear to me whether these would normally be processed with the dark water algorithm or not. This should be clarified and justified. It would be interesting to take some of these coastal sites and run an AERONET validation with them to see to what extent (if any) the RA algorithmic improvements presented in the paper improve the validation at these coastal sites. My expectation

C2160

is that the performance at these coastal sites should also improve. Additionally, I think it would be good to add a table giving island site names/locations and number of matchups, so we can see how these 954 cases are spread. I realise these are mapped in Figure 2, but it doesn't show the distribution of matchups between sites.

P7846, lines 12-17: I have a problem with this. I understand the authors want to stratify by chlorophyll concentration as a proxy for strength of underlight. But chlorophyll concentration varies by orders of magnitude, and the uncertainties on the chlorophyll retrievals can be (I think) something like 30%, potentially higher for Case II waters (where retrieved chlorophyll concentrations are often higher). The dependence of sensor-observed reflectance on underlight is also nonlinear. Given this, the split at chlorophyll thresholds of 0.25 and 0.5 mg m⁻³ is arbitrary. What is the justification for these thresholds? It seems like most of the data are in the lowest bound (not surprising as open ocean chlorophyll is often less than 0.3 mg m⁻³ except in coastal regions, and can be persistently 0.1 mg m⁻³ or less for large parts of the oceans). So given all the other factors influencing retrieval uncertainty, I don't know that the small number of high-C cases can really tell you anything about errors linked to underlight, if there is even any change in error characteristics. I would prefer a more objective/justified choice of thresholds here. If the authors just want a low/mid/high classification, why not just take the terciles of the chlorophyll concentration across the cases and report what they are? Or else why not calculate the underlight contribution to measured reflectance for each case (I suppose this would involve some weighting for different cameras) and stratify by that? That way the authors would be more directly looking at the underlight contribution, rather than using chlorophyll to guess where it 'might' be important. This stratification is also not really used much in the analysis, so perhaps it could be omitted entirely.

P7846, line 14: MODIS, SeaWiFS, MERIS acronyms should be defined on first use. I

C2161

note that MODIS at least is mentioned elsewhere in the paper, but I think this is the first instance.

P7847, lines 15-21: Table 5 appears to be mentioned before Tables 3 and 4. Figures 10 and 11 appear to be mentioned before Figures 5, 6, 7, 8, and 9. Following publication conventions, these should be reordered so they appear in order and proximity to where they are first discussed. I think it's fine that these contain some results that haven't been introduced yet.

Section 3

P7848, line 7: Do the authors mean that the reflectances in *each*, rather than *all*, available cameras are averaged? As if I understand correctly they end up with one reflectance per camera per band (i.e. up to 9×2=18 measurements) rather than one reflectance per band (i.e. up to 2 measurements).

P7849, line 9: See previous comment on this being a good place to re-evaluate the assumption that blue/green band ocean surface reflectance is not modelled well enough to use these channels in the RA.

P7849, line 16: What is the uncertainty on these data?

P7849, line 28: Probably best to be specific and say 'effective whitecap albedo' rather than just 'albedo'.

P7850, line 4: If I read Figure 5 correctly, it does not 'quantify the channel-by-channel

C2162

differences' from improving the surface model as the authors state here, but rather just shows the biases when the new surface model is used. Either the text or figure should be corrected.

P7851, lines 2-5: If I understand what is being said here, I think this could be rephrased to be clearer. Maybe something like the sensitivity to n_r is small such that its retrieval would be more sensitive to compensating errors in other parameters (e.g. AOD) rather than n_r itself?

P7851, lines 22-23 (and elsewhere): As the true refractive index is unknown in general, I think it would be better to say increasing/decreasing rather than overestimating/underestimation.

P7851, lines 24-25: Is this $1.28 \mu\text{m}$ single-particle radius, effective radius, number mean radius, volume mean radius, or what? I understand this is one of the components supposed to represent coarse marine aerosol; are similar results obtained for the $2.8 \mu\text{m}$ particle the authors mention? How much do these particles contribute in a typical over-water retrieval (i.e. does this affect results most of the time, or only rarely/as a small fraction of the AOD)?

P7852, lines 10-11: The term r_e should be defined here (I think this is first use), and units should be given for these numbers (I assume μm). Also, it would be good to state what these aerosol components represent and why they were added.

P7852, lines 20-22: Does this mean that when a cirrus particle is selected, there is no aerosol at all (i.e. 100% of the retrieved AOD is actually cirrus optical depth)? Also, it is mentioned that additional aerosol layers can't be added in the current

C2163

RA code; in that case, what are the vertical profiles of aerosols used? My guess is that for the wavelengths used in the dark water RA, aerosol vertical distribution will not be very important in the majority of cases, due to the low Rayleigh optical depth.

P7853, line 8: Are biases always positive? If not, I would suggest removing the word 'high'.

P7853, lines 21-23: What is the uncertainty estimate in the SA? Is it also 5%? If so, I would point this out here, instead of saying it is set arbitrarily in the RA. Presumably there was some evidence to support the initial use of this value.

P7854, line 6: It is probably worth mentioning that the 68th percentile is of interest as (for a Gaussian distribution) this represents one standard deviation of the data; some readers may not be aware of this.

P7854, lines 6-12: So this term represents the aggregate of measurement and forward model uncertainty. It would be interesting to know if this has any geometric or AOD-dependence. From Figure 8 it looks like geometric dependence may be small, as the different cameras have generally similar values for a given band. So what do the authors think is the main cause for the large bars in Figure 8 (I assume this is the central 50% of the data as per Figure 3, is that right)? Does it depend on AOD, solar angle, or something else? Or do the authors not find significant correlation with any other parameter?

P7854, lines 17-19: The authors provide here alternative total calibration/forward model uncertainty estimates for all four MISR bands. But is it still correct that only two (red and nIR) are actually used in the retrieval and solution-finding here? Due to

C2164

the component from forward-modelling the atmosphere/ocean system it may be that the blue/green wavelength values will not be appropriate for other (e.g. over-land) applications, as the authors note later on lines 24-25, and so presenting them without some caveats might cause others to use these numbers inappropriately (i.e. why calculate and present these numbers if you are not using them).

P7854, lines 19-23: This goes back to my previous question of why there is a weighting variable w in Equation 1, which doesn't appear to be discussed or used, as ρ_{err} is providing the weighting mechanism instead.

Section 4

General: I think here and elsewhere it would be good to be clearer about exactly what is done in terms of comparing to AERONET. As I understand it, there are potentially multiple aerosol mixtures which meet the acceptance criteria for a given MISR retrieval. So, in the various plots and discussions throughout, is only the 'best-fitting' mixture chosen, the mean of all accepted mixtures (this is what I infer is done in the SA at least from P7856, line 27-P7857, line 2), the median, some weighted average, or what? How much difference does this make to the reported AOD/ANG? How is this affected by the precise combination of various aerosol components into the different aerosol mixtures used (e.g. is one component 'rare' in the mixtures and so comparatively under-represented if solutions are averaged equally)? What should a theoretical data user do in this situation? How about the effects on other derived quantities, e.g. single scatter albedo? It isn't always clear what is done in the analysis, how the choice is justified, or what 'should' be done if/when these improvements are eventually transferred to the MISR SA.

C2165

P7855, line 13: I don't know that the word 'tiny' is needed/justified here.

P7855, lines 19-21: It would be worth mentioning here how these calibration biases change the statistics of the AOD comparison, i.e. you reduce the ANG biases but is there any corresponding cost or improvement to AOD validation?

P7856, lines 4-9: Was this new criterion determined empirically or is there some theoretical justification? If one assumes the uncertainties are well-characterised, then the statistics of the χ^2 distribution itself can be used to determine a theoretical cutoff of χ_{abs}^2 to use, for varying levels of confidence that the measurements and retrieved state are consistent with each other.

P7857, line 26: Given that most of the matchups are for fairly low AOD, what do these RMSE decreases equate to in terms of AOD? Is absolute or relative improvement the same across the range of AOD, or does one regime benefit more than others?

P7858, lines 8-13: Least-squares linear regression slopes are not useful/appropriate for this type of AOD comparison, in my view, for several reasons. The data are unevenly distributed in AOD space (AOD distributions are typically lognormal, not uniform), AOD error distributions are truncated in low-AOD cases (as negative AOD is not permitted), and AOD error itself is a function of AOD. These cause issues for naïve linear least-squares fits. Biases in high-AOD conditions may also be dependent on aerosol type, in which case it is not meaningful to talk about a global-average slope because it is so context-dependent (although I am not sure how strong type-dependence of bias is for the MISR RA; perhaps this is something else the authors could address, if they have sufficient data to make statements of this nature). For these reasons I would either not mention or not dwell on linear least-squares regression results. Some

C2166

of these issues can be ameliorated in global-average plots of this type by using more advanced linear fitting techniques, although others (e.g. opposing type-dependent biases) cannot.

Page 7858, line 27: See above; I would not bother discussing AOD regression slope here.

Conclusions (and some other suggestions)

This provides a good summary of what was done, but I think more emphasis should be given to the final paragraph here. As mentioned earlier, the RA is distinct from the SA in that the RA is used for case studies while the SA is the only product which has been processed on a large scale and is available to the public. The authors have done a good job of illustrating the advancements of the RA over the previous RA and current SA for these over-water cases. I believe that that MISR data have been under-used, for various reasons, compared to other satellite sensors (e.g. MODIS), but studies like this illustrate MISR's capabilities well. So one of the burning questions is which of these updates from the RA will be ported over to the SA, and when that is likely to happen. I understand that this will require effort on the JPL side of things as well. It would be good if the authors could make some statements about this.

On a related topic: the authors' analysis here focusses on select AERONET sites and MAN matchups. There aren't any larger-scale retrieval results shown, so it is hard to visualise the 'bigger picture'. I think it would be very valuable if the authors could undertake some more processing to show this. One suggestion, as mentioned earlier, would be to apply these corrections to some coastal AERONET sites in a few locations and provide tables (or plots similar to Figure 11) showing the SA/updated RA performance.

C2167

Just a few sites characteristic of different regimes (e.g. marine, dust, polluted) would suffice. There are plenty of long-term coastal AERONET sites in many areas where the existing island sites and MAN provide limited/no matches with MISR. This will be a good independent test of the improvements, and would help establish likely benefits of these improvements to the RA, for eventual incorporation into the SA, on a larger scale.

The second suggestion is to process some period of time (perhaps a month, but if that is too much, even a week would be better than nothing) globally of RA over-water data. Then show some global maps of AOD from the SA and updated RA. That would help visualise exactly how and where the improvements would be likely to affect MISR data. This is doubly helpful because, if there are areas where the SA and RA are very different, it tells us that there might be issues with current analyses based on SA data, while if there are areas where the SA and RA are very similar, it tells us that we can proceed with perhaps more confidence in SA data than we might otherwise.

Hopefully by posting this review with quite a long time period for the Open Discussion remaining, there will be the opportunity to process a larger quantity of MISR data with the RA without delaying the paper's revision too much. Ideally the authors will be able to include both of these suggested additions.

An advantage of MISR over many other Earth-orbiting measurements is that its multiangle measurements provide greater constraints on aerosol properties beyond AOD and ANG than many single-viewing sensors, such that in some conditions inferences about single scatter albedo, particle size, and particle sphericity can be made. AERONET data are, in some situations, also able to provide this information. I acknowledge that for the majority of the MISR/AERONET matchups here, this information may not be available. However I think it would be helpful if the authors could make some statements about how the other optical/microphysical aerosol

C2168

properties retrieved by MISR change when going from the SA to this version of the RA, and whether those changes make sense given our knowledge of regional/seasonal variations in aerosol properties. This could be accomplished by comparing against AERONET inversions for those cases where they are available, and/or by just showing the regional/seasonal dependence of some of these properties at selected sites. Even without case-by-case validation data, we might be able to see if they are noisy or stable, and if they make intuitive sense given our knowledge of the site in question. This will also feed back into my previous comment about the averaging strategy for acceptable mixtures. I think an analysis of this type would be of benefit to the paper as a way to further illustrate some advantages of multiangle sensors like MISR (as only looking at AOD/ANG is, in my view, selling MISR short).

Supplement

The Supplement is not really referenced by the main manuscript and, in my view, does not add much. Many of the figures are hard to interpret and find the take-home messages without additional explanatory text. I suggest the authors remove the Supplement. If they feel there is key information here (e.g. some of the refractive index information referenced on P7852, or discussion of MLM on P7853), it should be integrated into the main manuscript, or the Supplement should be expanded so it can be understood on its own. However my general view is that the contents of the Supplement are not particularly necessary for this study (although they are certainly valuable for the MISR team's development) because the most important results are contained within the paper, and the general AMT readership do not have access to the RA data anyway.

C2169

Tables and Figures

Table 2: As mentioned before I don't think this chlorophyll stratification is optimal, and perhaps not necessary.

Figure 4: Presumably the colour scale is fractional density of data? This should be stated in the caption. Also, I suspect that when viewed at 100% size in final journal layout, the annotation here will be too small to read. Perhaps the font size could be increased, and some of the information that is given elsewhere or not too important (e.g. number of coincidences, slope) removed.

Figure 5: I think the words 'under-light bias' should be removed as they don't add anything and could be confusing. What I think is shown here is the bias *after* underlight (and whitecap) updates are made, which is not what the phrase implies. I would just delete the additional text in the middle of the plot. Also, I assume the grey horizontal lines here (and in e.g. Figure 3) are the AERONET direct-Sun AOD uncertainty of ± 0.01 ; this should be stated somewhere in the captions.

Figure 6: I would drop the word 'theoretical' from the caption as in my view it is kind of empirical (even if it is synthetic data) and I don't think the word adds anything. Additionally, since results appear to be consistent going East-West across the orbit, this might be clearer to present as a series of line graphs (for different AODs and latitudes; I assume the changes with latitude are a result of changes in solar zenith/azimuth angles, so it might be better to frame it that way). If the authors decide to retain the maps, then the orbit-average change in AOD and χ_{abs}^2 should be removed. I don't think that orbit-averaged values are meaningful (as the Figure shows, these things change fairly systematically along the orbit and across-track) and in any case the symbol ϵ was used for an unrelated quantity in Equation 3, so repetition of the symbol is confusing.

C2170

Figure 7: The same basic comments as for Figure 6 apply here.

Figure 8: The left-hand side of the text at the foot of this Figure is slightly chopped off.

Figure 9: Comments for Figure 4 about annotation size also apply here.

Figure 10: Comments for Figure 4 about annotation size also apply to panels (b,c). Also, I would consider splitting Figure 10(a) from (b,c) for brevity of caption, if nothing else. The sentence about the log x-axis for Figure 10(a) is not necessary and can be removed for brevity.

Figure 11: The sentence about the log x-axis is not necessary and can be removed for brevity. I also think the vertical range (at least for the blue band) should be increased here, because it looks like the 68th percentiles for the two SA lines for the top AOD bin are off the scale.

Interactive comment on Atmos. Meas. Tech. Discuss., 7, 7837, 2014.