

The authors would like to thank Andrew Sayer for his very thorough review. We have made many changes as a result.

Abstract

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

Corrected.

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

Done.

P7838 line 12: Presumably spectral ‘calibration’ uncertainty.

In this case the uncertainty is not just calibration uncertainty, so we leave the statement as is. This uncertainty includes AERONET uncertainty, surface modeling uncertainty, aerosol model uncertainty, RT, etc.

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.

Corrected.

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.

Corrected.

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.

Thank you for pointing this out, this has been corrected.

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 If you search the literature you often see incorrect/outdated citations to datasets so I think it's best to nip this in the bud.

We have added a statement in the introduction and conclusion stating that the RA is only a research algorithm, and it is not practical to generate global datasets with this code.

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.

This has been removed.

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 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?

As implemented, the code runs the dark water portion of the algorithm if the SA runs the dark water algorithm and the sub-region used has a SurfaceFeatureID ≥ 5 (dark water). We have added this statement to section 2. We could run many coastal sites with the dark water algorithm, but there are other issues as well (such as coastal run-off, differences in aerosol amount and type between inland AERONET stations and just off the coast, coastal biological activity, and aerosol variability) that would hamper our validation efforts. Again, these sites are generally dark water (as are large, deep lakes and seas). The question would be worth re-examining if an upgraded algorithm were run globally, but this is not expected in the near-future.

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.

We have removed A_0 and moved the sentence to section 3.2.

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

Corrected.

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'.

We have defined mixtures in the introduction.

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 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.

We agree with including cloud contamination as the fourth factor, and have done this. We attempt to explain that the uncertainty due to structural aspects of the algorithm (lack of polarization, optical model errors, vertical distribution, etc.) is included in our uncertainty estimate, as is the uncertainty in AERONET/MAN data itself.

P7842, lines 9-11: Is there a reference for the scatterometer data? What is the un-certainty 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.

The data comes from Quikscat until Nov. 2009, then SSM/I. Unfortunately, we found out that both the first-look and final wind data found within the MISR aerosol file use monthly wind speeds, which means both the SA and RA used monthly winds. Because of this, we have re-run the RA cases with 6 hourly Cross-Calibrated Multi-Platform (CCMP) Ocean Surface Wind Velocity Data. Because this data is only available through 2011, we will use the MISR monthly winds when this data is unavailable.

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.

You are correct that we use a “brute-force” method, but we compute χ^2 for each point on our fine AOD grid. We have clarified this in the text.

P7843, line 12: I think the denominator of the upper sum should be typeset as

p_{err2} rather than p_{2err} , 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. Thank you for pointing this out. We have corrected this.

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 p_{err} (which is effectively a weight on each measurement)?

The weights are camera-dependent and are currently set to $1/\mu$. If we were to include this weight in p_{err} , p_{err} would no longer represent the band/camera specific uncertainty. We have now included the value of w just after equation 1.

P7843, lines 21-25: These additional metrics should be explained more clearly and defined explicitly. I infer χ_{2geom} and χ_{2spec} 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 $\chi_{2maxdev}$ simply the largest individual (band/camera) contribution to χ_{2abs} ? 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.

Considering we are not using the metrics other than abs and maxdev in this paper, we will remove references to them. Maxdev is simply the largest un-weighted contribution to abs, but the equation for it has been added.

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?

We had already looked at including the blue and green bands after including the ocean surface model. Adding the blue, even with minimal weighting, severely degrades the results. Including the green (even if only for mixture selection and not for AOD retrieval) does not improve the skill of the algorithm, and given the uncertainty in ocean-color products, we think it makes sense to not include either. Given the systematic nature of the errors in the blue, we are very confident that we are seeing surface modeling errors. However, we added the words “among other factors” to the surface modeling error description as the reason for only making use of the red and NIR wavelengths for our dark water retrievals.

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

As implemented, the code runs the dark water portion of the algorithm if the SA runs the dark water algorithm and the sub-region used has a SurfaceFeatureID ≥ 5 (dark water). We have added this statement to section 2.

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

In the sentence prior we defined the residual without an absolute value sign.

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

We have changed to “much brighter than the expected surface brightness”.

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.

We are using AERONET level 2 version 2. We have added the suggested references.

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.

We have moved this to the introduction.

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).

The authors meant that the ships move, whereas AERONET sites do not, the platform is less stable, and there is more room for operator error. A ship can easily move more than 1 full retrieval region (17.6km) in an hour's time. We have clarified this in the text.

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 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. These sites would be processed by the dark water algorithm, but there are other reasons why they are less desirable to use for validation. AERONET may not see the same aerosol type/amount over land, compared to what MISR sees over water. In addition, run-off can be a major problem near coastlines, and this is an algorithm about dark water, not coastal regions. We replaced Table 2 with an table showing the distribution of the various AERONET sites, with the number of coincidences included.

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.

We will eliminate this stratification entirely. Thank you for your input here.

P7846, line 14: MODIS, SeaWiFS, MERIS acronyms should be defined on first use. I note that MODIS at least is mentioned elsewhere in the paper, but I think this is the first instance.

Corrected.

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.

We have changed the table order, but the figure order would be very confusing if we were to change that because of one brief mention to Figures 10 and 11.

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 $9 \times 2 = 18$ measurements) rather than one reflectance per band (i.e. up to 2 measurements).

In this case, all 18 measurements are averaged, such that we would have 16×16 values. Then the darkest pixel is selected from those 256 values.

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.

It is clear from the previous data that the blue is not suitable for use. We have added a statement to this effect at the end of section 3.2

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

The CHL data has an rms error of 0.3993, and the CDM data has an rms error of 0.3720. However, these data include many cases where CHL is >1 , and a few cases of $\text{CHL} > 10$. The files containing the CHL and CDOM data also have a CDM error and CHL error dataset [Maritorena et al., 2010]. We have added a sentence pointing the readers to this reference.

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

Done.

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

You are correct. We have edited the text.

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?

We believe our current sentence states this.

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.

In this case, we were comparing model-to-model results, so we know n_r . Also, we want to emphasize that we are not randomly choosing n_r values (to us, increasing/decreasing suggests this). The reason we changed the values of n_r for our particles is that we believe the n_r values (found in the table, decreasing with r_e) better represents these particle in nature.

P7851, lines 24-25: Is this 1.28 μm 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 μm 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)?

This is effective radius. The results presented for the 74-mixture set assume a 2.8 micron coarse mode particle (with a n_r of 1.45), whereas the results presented for the 774 mixture set assume a 1.28 micron coarse mode particle (with a n_r of 1.37). The supplemental material is cited at the end of the section, as it contains the n_r sensitivity study. As for how much particles contribute to typical retrievals, *Kahn and Gattley* [2014] wrote a paper on this.

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.

Corrected. We have added this information.

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 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.

Correct. The vertical profiles are not listed because they vary with mixture. The vertical profiles for the mixtures vary with mixture. Mixtures with more dust have their peaks higher (3-6 km) than mixtures with more spherical components (0-2 km). Cirrus mixture profile peaks are > 10km high (mostly above the Rayleigh layer). Although you may be correct about the vertical distribution, we will refrain from using cirrus-mixed mixtures. When cirrus is present, it can easily overwhelm the aerosol AOD in the TOA signal, and given the limitations of the current RT code, we model it as the sole component in this study.

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

For the work we did, and over dark water in general (*Kahn et al.*, 2010), the errors are almost always positive.

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.

Yes, it is also 5%. We have removed the word arbitrary.

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.

Done.

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?

You are correct that the error bars represent the central 50 % of the data. We believe part (up to ~15%) of the issue may be a calibration drift. Clouds appear to be another part of the issue (up to 25% in the blue, less in the longer wavelengths, based on some figures we have looked at with our data). A large portion of the remaining errors are likely due to "ghosting" issues.

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 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).

Good point, we have added the caveat.

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 p_{err} is providing the weighting mechanism instead.

We have added that w is $1/\mu$ (In section 2.1), but this is not used for estimating the uncertainty.

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.

The mean of all accepted mixtures is used for AOD in the RA for this study, as in the SA. Using the mean instead of the best fitting mixtures tends to give us lower RMSEs, in part because the actual particle optical properties don't always match the options in the algorithm. This represents an uncertainty in the retrieval. In general, as AOD increases, fewer mixtures are selected by the algorithm, so the fact that we average these parameters becomes less important. At very low AOD, we do not have sensitivity to particle properties, so the best-fitting mixture (or the mean) might not represent reality any better than an arbitrary choice. Under these circumstances, the information content is lacking in the measured reflectances, many mixtures pass, and we are experimenting with using a chemical transport model to weight to more likely among the passing mixtures [*Li, Kahn, et al.*, JGR submitted]. Your point about rare components is well taken, but looking at our mixture list, it is not an issue (except for cirrus). Validating the particle microphysical properties, except ANG, is beyond the scope of the current paper, but the way the SA presents these parameters is very similar to the way we currently do.

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

We have removed "tiny".

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?

We have added that this adjustment does affect the AOD agreement statistics.

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^2_{abs} to use, for varying levels of confidence that the measurements and retrieved state are consistent with each other.

This is determined empirically. The fractional uncertainty of the data (especially the NIR) depends on the magnitude of the radiances. At higher AOD for instance, the fractional uncertainty in the NIR is much lower than at low AOD. To the extent that the AERONET data represent ground truth for AOD, the empirical assessment is the best we know how to do.

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?

Although the numbers are not shown in most plots, you can see the answer in Figure 11. Most of the improvement is found at shorter wavelengths and lower AOD, but there is improvement in almost every AOD bin for all four wavelengths.

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 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.

We agree, though this is commonly done in AOD studies, so we removed information about the slopes for AOD. But given that ANG does not have cut-offs at zero and should be more evenly distributed, we are leaving the ANG slope discussion alone.

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

Done

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.

We have added a statement clarifying that these data are not publicly available in the conclusion. As for what improvements will be ported to the SA, we would hope that changes to nr (which are easy to address), under-light (which is again not difficult to address), and hopefully an updated mixture list will be included in the next version, but we are in no position to determine the process at JPL, and cannot say what they will actually implement, or on what schedule.

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 under-take 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. 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.

In the paper, we already consider a wide distribution of AERONET + MAN coincidences, covering a range of dark water situations (Fig. 2). Coastal sites provide different challenges, as the water surface might not be dark, and there might be significant differences between the aerosol over coastal water and nearly adjacent land. The component of the RA studied in the current paper deals only with such dark water situations. Bright-surface retrievals are grounds for a separate study.

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.

Note that the current SA has been validated in separate studies, so the uncertainties in that product are documented against AERONET and some case studies, as you know. Unfortunately, it would take too long to process even a week's worth of data with a sufficiently rich mixture climatology. The algorithm is not currently programmed to efficiently deal with multiple orbits at a time. Although we are not supported to generate anything like an operational algorithm, we are always looking to improve speed and accuracy though, so hopefully this will be feasible at some point in the future.

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 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).

The current paper is already quite long, and we elected to focus on AOD and ANG, for which we have substantial ground-truth from AERONET. *Kahn and Gaitley* [2014, JGR submitted] wrote an entire paper validating the particle properties from the SA; for the RA, this would be a separate, though useful, study.

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.

The Supplemental material is important for demonstrating the need to include realistic real refractive indices, and to establish the limitations of the MLM approach over broader parameter spaces than are covered in the main paper. We use the results of these analyses in the main paper, and have expanded the Supplemental text to explain what was done more fully.

Tables and Figures

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

We agree, so this has been eliminated.

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. You are correct about fractional data density, and we now state it explicitly. The key textual information is found in Table 5, so we have removed the text from the figure.

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.

The plot title has been removed, as has the text in the middle of the plot. The grey lines are there for reference, and are now explained in the caption.

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_{2abs} 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. We removed “theoretical”. For these plots, we only perform one retrieval per block, so that is why there is no change East-West. We have made this clearer in the figure caption, and removed the orbit-averaged values.

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

We removed “theoretical”. For these plots, we only perform one retrieval per block, so that is why there is no change East-West. We have made this clearer in the figure caption, and removed the orbit-averaged values.

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

This is fixed.

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

We made the text larger.

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.

We removed the text from b and c, and removed the sentence about the log x-axis.

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.

Corrected.