

# ***Interactive comment on “Updated MISR Dark-Water Research Aerosol Retrieval Algorithm – Part 1: Empirical Calibration Corrections and Coupled 1.1 km Ocean-Surface Chlorophyll-a Retrievals” by James A. Limbacher and Ralph A. Kahn***

**Anonymous Referee #2**

Received and published: 6 January 2017

Overall, I think this is a good paper that merits publication. The authors address the capability of MISR to account for ocean reflectance by parameterizing Chlorophyll-a concentration in the water. Considering the long history of MISR, this was overdue for an investigation.

However, there are some organizational, conceptual and methodological problems in the paper that mask its merits. I think addressing these issues will make for a paper this is more useful to the community (and citable).

C1

One of my main problems is with how the paper is organized, and the relative weight given to describing tweaks to MISR calibration and the incorporation of Chl-a in the Research Algorithm (RA). These are two entirely separate issues, and honestly I think the audience for the latter is far larger than the former (especially considering the literature by the authors and others about calibration in the last few years). This paper feels more like a progress report of all activities performed in a period of time, which isn't the best way to present results to the community. I think this would be a far stronger paper if the calibration changes were either published elsewhere, or in an addendum or the supplementary material. At the very least, a clearer separation in the paper between topics related to calibration and the coupled Chl-a retrieval is needed.

I'm also concerned with the methods used to validate the Chl-a coupled retrieval. Comparisons with SeaBASS are an obvious way of doing things, but of course the number of match-ups available for comparison are limited. What bothers me is the 'validation' against MODIS results. As the authors note, the multi-angle MISR data have access to more information about aerosols, while MODIS has channels better suited for Chl-a retrievals. Differences between coupled algorithm MODIS and MISR results could be due to either instrument – MODIS is not necessarily a standard to which MISR should be held. What is the value, then, of a scatterplot of MODIS and MISR Chl-a results?

Continuing with a discussion coupled Chl-a retrievals, why is there no attempt to determine the uncertainty in such a result based on the components of the RA? SeaBASS data uncertainties are mentioned at one point, but why are those uncertainties are also not incorporated into any validation?

There are a number of statistical tools available to test the hypothesis that dataset A is identical to dataset B. These simple tools go beyond comparisons of correlation coefficients, etc. and actually state the confidence intervals for agreement, taking into account things such as sample size. A problematic example is in the abstract, where it is noted that MODIS has a higher correlation coefficient (0.91) than MISR (0.86). Given the small sample size, these differences are probably statistically insignificant.

C2

Please get an statistics textbook, learn some hypothesis testing techniques, and start applying them. I've also found the following publication to be useful even if it comes from outside our discipline: Altman, D. G. & Bland, J. M. (1983). Measurement in medicine: the analysis of method comparison studies. *The statistician*, 307–317.

I also wonder why the authors never tested the new RA coupled algorithm on synthetic data. Given the limited SeaBASS dataset, doesn't at least confirming the new algorithm can successfully operate on synthetic data have merit?

Presumably this is part one of a two (or more?) part series? I think how this paper fits in that series needs to be discussed.

#### Detailed comments

Page 1, line 19-21: I fully do not understand the first sentence of this paragraph – you are running the algorithm to validate the algorithm and somehow also analyze corresponding MODIS data? I think this becomes clearer later on in the paper, but at this point this serves more to confuse than illuminate

Page 1, line 21-23: Does it really make sense to compare a correlation coefficient of 0.91 to 0.86, especially for a small dataset? I think a better way of saying this is that they are statistically identical... which brings me again to the point that you should be using hypothesis testing.

Page 1, line 25: it's not clear at this point the meaning of looking only at  $\text{Chl} < 1.5$

Page 1-2, final abstract sentence: While I agree that this might be the case, I'm not sure you've demonstrated this in this paper, particularly for the value of joint MISR-MODIS retrievals.

Page 2, line 13: an appropriate continuation of the last sentence of this paragraph would be "assuming that ocean and aerosol signals at TOA don't co-vary."

Page 3, line 27: It would be good to give either a description or a reference for where

C3

the MISR SA wind data come from, and not assume all readers inherently know this.

Page 3, line 27: Why doesn't MISR use NCEP or other modeled products for sea surface pressure? Is that what the SA does, and you're now setting this to 1013.25 because of... mountains near the ocean?

Page 4, lines 2-4: I think separate weighting for aerosol and ocean components of a retrieval would be difficult to implement practically, as the appropriate weights are most likely scene dependent.

Page 5, lines 20-30: I'm really not a fan of arbitrary/empirical weighting algorithms that are presented as fact without any description of the logic behind the choices that were made and their expected significance.

Page 8, line 7: Please use a proper citation for that ATBD or relevant paper, not a website.

Page 9, line 2: "Spectral coefficients of variation" – this is imprecise wording. IS that the CoV for all of one channel for a patch, added up for all channels?

Page 9, lines 16-23: Does this means surface BRDF is assumed to be isotropic? Is that the case?. Also units should be given for the equation on line 17.

Page 10, lines 1-10: Isn't any long term linear trend determined in 3a? Or is this done a year at a time? It is unclear to me from your description.

Page 10, lines 22-24: While I'm sure you're familiar with the Bruegge paper, others may not be. This sentence is quite confusing on its own and needs a more detailed explanation.

Page 11, five starting bullets: all of these are arbitrary choices. It would be nice if you had an explanation for your logic in choosing them.

Page 11, five starting bullets: I'm still not following the logic of why MODIS data should be used to screen MISR results, if that is indeed what is happening (its not clear).

C4

Page 11, line 10: Here you say you flag out Chl > 1.5, but this is presented in the figures? Do the figures show flagged results removed, or not?

Page 11, line 16: what is the value of comparing the mean of MISR and MODIS data to SeaBASS?

Section 4.1: again, I think there are statistical tools you can use that do hypothesis tests that account for sample size. Then you can say in a more quantitative way that the amount of comparison data is “too few” (if it is).

Section 4.1: You mention in several points that collocated MISR data could improve MODIS retrievals of Chl-a. While I agree that this is probably the case, nothing in the analysis you've presented in this paper can demonstrate that.

Page 12, lines 1-3: Again, comparing MODIS and MISR doesn't necessarily indicate the correctness of the MISR algorithm. I could envision some situations (probably with low AOT) where MODIS would work better, and others where MISR would. I think if you had parsed Fig 7. to scenes with high and low AOT, or varying Chl-a, you could start to illuminate these issues.

Page 12, line 14-15: I agree with this

Page 13, line 11: “substantially better” Is this the case, or is this just a symptom of correlation coefficients, etc. calculated with far more cases than the SeaBASS comparison.

Page 13, line 19: Are coupled retrievals performed for MISR scenes where some cameras observe glint, or are they only computed when all cameras observe a scene? If it's the former, do these retrievals work as well since there are fewer angles? Presumably capability would be degraded – how would this affect the ability to fill in MODIS Terra data in glint?

Figure 1: I really had a hard time understanding this flowchart. Please try to make it more legible. Minimize text where you can – less is more.

C5

Figure 2: It would be nice to remind readers the actual wavelength of the blue and green channels

Figures 4-5: Nobody knows what “Camera Aa”, etc. means outside the MISR group. Please just state the viewing angle instead.

Supplementary material

Figure S1: Can you see the colored value of the AERONET sites? I can't.

Only here to you allude to a “part II” of this study. I can only assume that portion will have more details about the coupled RA. Which makes me think that there should really just be two papers, one about calibration, one about coupled retrievals.

All supplementary figures: I'm confused why there are consistently fewer MODIS retrievals in some of the scenes, can this be discussed?

---

Interactive comment on *Atmos. Meas. Tech. Discuss.*, doi:10.5194/amt-2016-360, 2016.

C6