Anonymous Referee #2

Received and published: 26 September 2017

Review of "New approach to the retrieval of AOD and its uncertainty from MISR observations over dark water" by M. Witek et al. for AMT

Synopsis: This paper describes a new method for retrieving AOD over water, using MISR observations. Specifically, instead of picking a retrieval solution based on the minimum cost ("best") fitting of lookup table versus observations, the new algorithm retrieves based on weighting the cost of each ensemble member. Instead of thresholds, the new retrieval is more dynamic, and appears to provide more accurate and more consistent results. Additionally, a new confidence index (known as ARCI) is proposed, which can help to screen the results. In this way, the uncertainty of the retrieval is quantified.

Assessment: This is a good paper, and should be published after minor/medium revision. The most obvious issue is that there is neither "validation" (comparison with ground-truth, e.g. AERONET) nor detailed comparisons with other datasets (e.g. MODIS on the same Terra platform). Based on my own experiences, I agree that the new results seem better (lower average AOD; fewer blunders, etc). However, a more skeptical reviewer needs some more proof including validation. I also wonder why the previous (≤ V22) retrievals had such a complicated chi-squared decision tree, when in fact it seems to be much simpler? The paper appears to be primarily about the advantages of the new ARCI/chi-sq metrics, which is fine. The issue becomes confused when discussing new aerosol model/mixtures, and much more confused when discussing 17.6 vs 4.4 km resolution. I recommend ONLY concentrating on the new fitting metrics here, because that is useful enough.

Re: We carefully considered including some form of external validation of the new approach (AODs and their pixel-level uncertainties) in this manuscript, but eventually decided the topic is challenging enough to deserve a separate study. Here we will try to briefly summarize our reasoning behind this decision. First, at the time of writing, only two months of V23 data were available, which did not provide enough comparison points against surface-based AERONET observations. At present, we have processed two years, 2014 and 2015, and obtained around 1300 collocations with AERONET. Note that we are constrained to Dark Water retrievals only, which limits the number of available AERONET locations. This number could be sufficient for AOD validation, but in our opinion it is still insufficient for a proper assessment of the reported pixel-level uncertainties. There is a range of topics that we would like to explore while assessing the MISR AOD uncertainty predictions:

- How do the spatial and temporal differences between MISR retrieval and AERONET observation influence agreement metrics?
- Is spatial variability in AOD uncertainty consistent with expectations?
- Is the AOD uncertainty dependent on specific retrieval parameters (e.g., viewing geometry, number of cameras used, ARCI parameter)?
- Is the AOD uncertainty affected by the proximity of clouds?
- How can we use information from other instruments (MODIS) to evaluate the AOD uncertainties?

These are just a few questions that we have already started investigating. In our view, a cursory evaluation within the scope of the present manuscript would have been unsatisfactory.

In this study we introduce the ARCI metric as a screening parameter and highlight its efficacy, but, in our view, this work is primarily about a new way of determining AODs and AOD uncertainties using the full information content available from the goodness-of-fit metrics. In

particular, this leads to a more plausible prediction of the AOD retrieval uncertainty, which we hope may prove useful in many aerosol modeling applications. An unwelcome side effect that we discovered after introducing this new approach was a relatively large number of high-AOD retrievals in areas that typically have low aerosol content, but at the same time are very cloudy. We concluded that these high-AOD retrievals were likely cloud contaminated due to imperfect cloud identification procedures in the MISR aerosol retrieval algorithm processing. In the V22 product, various thresholds on χ^2 metrics were able to eliminate many such erroneous AOD retrievals. In V23, the new ARCI metric is a useful alternative to the V22 thresholds. The transition to a finer horizontal resolution, from 17.6 km^2 to 4.4 km^2, fundamentally increases the number of cloud-contaminated retrievals because the retrievals are often performed closer to cloud edges, and some of the cloud screening that was effective at the coarser resolution was found to be ineffective at the finer resolution. We do not, however, discuss in this manuscript the impact of the finer resolution on the quality of retrieved AODs and AOD uncertainties. This will be a subject of a separate investigation.

Also, with the subject being the new ARCI/chi-sq metrics, I would be completely curious to see what these look like on the globe? (function of season, perhaps?)
Re: Yes, this is an interesting question that we will investigate in the near future. The paper's main focus is on the new methodology for deriving AODs and AOD uncertainties in the new V23 MISR aerosol product. Including additional analysis of ARCI would, in our view, diverge the manuscript from its main topic.

Writing: While the English writing is easy to read, there are issues of paragraph formatting (hanging vs indents). References are hard to read etc.
Re: We will format the references to be more transparent.

Specifics:
*P1L15: Why only allow AOD < 3.0? sometimes even higher?
Re: MISR aerosol look up table (LUT) only includes mid-visible AODs below or equal to 3.0. It is possible to extend this range to higher AODs, but to do so requires a significant change to the LUT and adversely impacts the processing time.

*P2L22: Suggest using the term "confidence" rather than "quality", as the MODIS retrieval can't measure quality until performing validation. Confidence refers to how well the algorithm marched through its logic steps (enough pixels? Good enough fitting? Etc).
Re: Yes, we generally agree with this statement, but in this case we refer to the Quality Assurance (QA) metric specified in the MODIS product. In Levy et al. (2013) on page 2990 we read: "However, the are major changes to how data "confidence" or Quality Assurance (QA) is assigned (Hubanks, 2012)." As we refer to a flag in our sentence, we think the phrase "retrieval quality assurance flags" is appropriate.

*P2L29: Suggest adding where these uncertainties would be useful, especially in applications of data assimilation/forecasting etc.
Re: We modified the last sentence in this paragraph to read:
"While such metrics are very valuable, they comprise only crude proxies for pixel-level uncertainties and, therefore, have limited quantitative utility in applications such as aerosol forecasting and data assimilation."

*P2L35: Note that the MODIS retrieval (and I think others) do not validate in terms of ±MAX(a, b x AOD), but rather as ±(a+b x AOD).
Re: Yes, most satellite instruments retrieving AODs report their error envelopes as ±(a+b×AOD).

MISR defines the error envelop in a slightly different manner. We changed the sentence to read: "Taking the general form of ±(a+b×AOD) (or max[±a, ±(b×AOD)]), where a and b are empirically determined constants…".

*P3L8: Ensemble approach. YES! We have more computer power, I agree! Note that the MODIS over-ocean retrieval does a poor-man's ensemble.
Re: Agreed.

*P4L12: Are you reviewing the old algorithm (v22) or the new one (V23)? Or is everything common to both?
Re: We modified the sentence to read:
"Here some key elements of the V22 algorithm relevant to the new approach are reviewed.

*P4L15-17: This sentence is a run-on and confusing
Re: We rearranged this sentence to read:
"The problem of retrieving aerosol properties over large water bodies, such as oceans, seas, or deep lakes, is greatly simplified by the fact that reflectance from such surfaces is uniform and that such deep-water bodies are essentially black at red and near-infrared (NIR) wavelengths."

*P4L17: Not sure what the sentence about 1-D RT means.
Re: It is a general statement regarding the physical principle of AOD retrieval over dark water.

*P4L36: So this more comphrensive model set is not used for V23, correct?
Re: Correct, V23 includes the same set of mixtures (and same LUT) as V22.

*P6L3: What happens to fitting error if AOD is near zero? Very low signal.
Re: In Eq. 2 for χ^2, the signal difference (ρ_{MISR}-ρ_{model}) is divided by σ^2_{abs}, defined in the text (P5L16), which takes into account the signal magnitude.

*P6L28: This sentence is a run-on.
Re: Agreed. We rearranged this sentence to read:
"The empirical thresholds in goodness-of-fit parameters in the V22 MISR dark water aerosol retrieval algorithm are used to select successful aerosol mixtures. This affects the frequency of retrieval success as well as the resulting AODs, AOD uncertainties, and aerosol properties."

*P6L31: What is a "blunder"? Is this a retrieval by mistake? No retrieval when should be? One with a big error? Do you really want to screen all "outliers"?
Re: A retrieval "blunder" is a retrieval with very high AOD that is untrustworthy and possibly affected by cloud contamination. Reasons other than cloud contamination are also possible. Ideally, cloud identification procedures should be able to eliminate all cloud-contaminated pixels so that an aerosol retrieval is not performed. However, most satellite instruments suffer to some extent from erroneous cloud identification, in which case cloudy pixels are used in aerosol retrievals. This results in clouds being retrieved as aerosols with unreasonably high AODs.

*P7L3: Does Fig. 1 represent a particular date/time/case? I know it is discussed further in a future section, but it's confusing here. At least mention that it will be discussed more. I however, like the visualization. What happens in case of bigger (or smaller) AOD? Will the spreads be smaller or larger?
Re: This is a randomly selected case. We added appropriate clarification in the text:

"The key elements of the new method are visualized in Figure 1 using actual MISR data from a randomly selected case."
Generally yes, the spread, and the uncertainty, depends on the retrieved AOD, which we visualize in Fig. 6.

*P8 last paragraph: I am getting confused because paper is discussing TWO upgrades. (A) The ARCI/chi-sq stuff, and also the (B) Spatial resolution (17.6 to 4.4 km). I think you need to concentrate on only (A).
Re: The increased resolution of the retrieval is not an upgrade that we are concentrating on in this manuscript. The processing pathway is exactly the same in both the 17.6 and 4.4 km retrievals, except that the 4.4 km retrieval covers a smaller area. In fact, the MISR Dark Water algorithm at either resolution selects only one 1.1 km pixel, which is then used to perform an aerosol retrieval. This one pixel in V22 is assumed to represent an area of 17.6 x 17.6 km, whereas in V23 it represents an area that is 16 times smaller (4.4 x 4.4 km). This is why retrievals are often performed closer to cloud edges.

*P9L16: Why is low ARCI related to cloud contamination? It is definitely one reason. Could there be confusion between small ice particles and dust particles, and somehow derive a large ARCI?
Re: In our analysis we observed a relationship between the prevalence of high-AOD retrievals and low ARCI. These high-AOD retrievals are in areas that climatologically have very low aerosol burdens, but are characterized by high cloud coverage. Cloud contamination in the MISR retrieval appears to be the most plausible explanation for such high-AOD results.
There are certain conditions when the MISR retrieval algorithm identifies thin cirrus clouds as non-spherical mineral dust mixtures. This was documented in a study by Kalashnikova et al. (2013) and manifests itself as bands of aerosol nonsphericity over high latitude oceans (e.g., the Southern Ocean, Northern Atlantic) that shift with the seasons. This is clearly an issue of cloud contamination. Those retrievals, however, tend to have low ARCI, and the new screening approach based on the ARCI threshold is able to eliminate them.

*P9L20 (and Fig 3). Hard to see, because panels (b) and (d) have different y-axis scales and they are not in terms of %. To me, it looks as if there are much fewer retrievals in panel (d) versus (b). Also, why the wiggles in (b)?
Re: The maximum values in Fig. 3(b) and (d) are different, but the scale is linear in both cases. We concentrate on the trend in retrieval count, rather than on the absolute values, which depend on the spacing of the ARCI and min(χ^2) parameters. In this particular case, we used 200 intervals for min(χ^2) (range from 0 to 5), and 290 intervals for ARCI (range from 0.013 to 0.4).
We do see certain clustering around specific min(χ^2) values in our dataset, which gives rise to small wiggles seen in Fig. 3b. This is probably related to the finite AOD gridding of our LUT, which is 0.025 throughout most of the AOD range. We plan to investigate this feature in greater detail in the future. Furthermore, the wiggles in Fig. 3b become apparent only because of very fine sampling of the min(χ^2) space, which is 0.025 in this case.

*P9L32: Is there a chance you are throwing out "good" aerosol data? Maybe you can show some AOD imagery (on a map) over-plotted on the suspected clouds?
Re: We have not looked at particular cases or extensively investigated specific regimes in Fig. 4a. However, motivated by your comment we looked at the origin of this particular group of high-AOD retrievals with min(χ^2) around 0.2 and ARCI around 0.1. This turns out to be about ~410 retrievals coming from one orbit in 18th January 2007 (orbit 37689). To our surprise, these retrievals are south of the Ivory Coast, Africa. The figure below shows unscreened AODs from MISR V23. There are some scattered clouds in the scene but they are not related to the patches of

high-AOD (>2.0) retrievals. The aerosol background is high with AODs exceeding 0.5. The second figure shows MISR equivalent reflectances from the red wavelength for the same scene. This is to show that the "plume" of high-AOD in the first figure does not correspond to the higher radiances measured by the instrument. The visible imagery from MODIS also corroborates the finding that there is no substantially thicker aerosol plume is this area. This strongly suggests that the retrieved AODs in this region are retrieval artifacts, likely related to the mismatch in assumed aerosol properties between the current MISR LUT and reality, which may be a smoke and dust mixture not contained in the current MISR LUT. The current ARCI threshold screens out these "poor" retrievals.

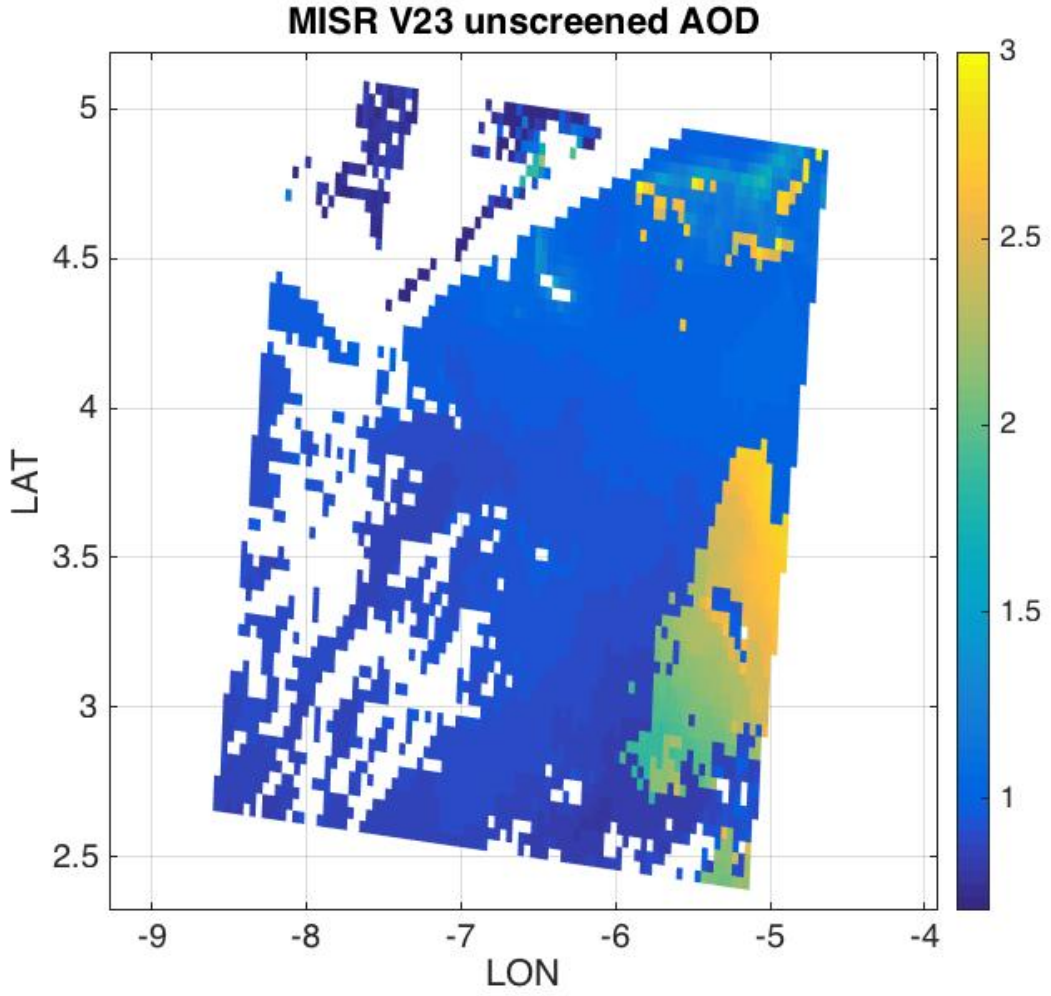

Figure 1 MISR V23 unscreened AOD from orbit 37689, blocks 87-88, time: January 18, 2007, 10:58 UTC. The high-AOD retrievals in the center right of the image have low ARCI and are therefore screened out in the final product. These retrievals, however, have min(χ^2_{abs}) values below 2.0 and therefore would have passed the in the previous V22 algorithm.

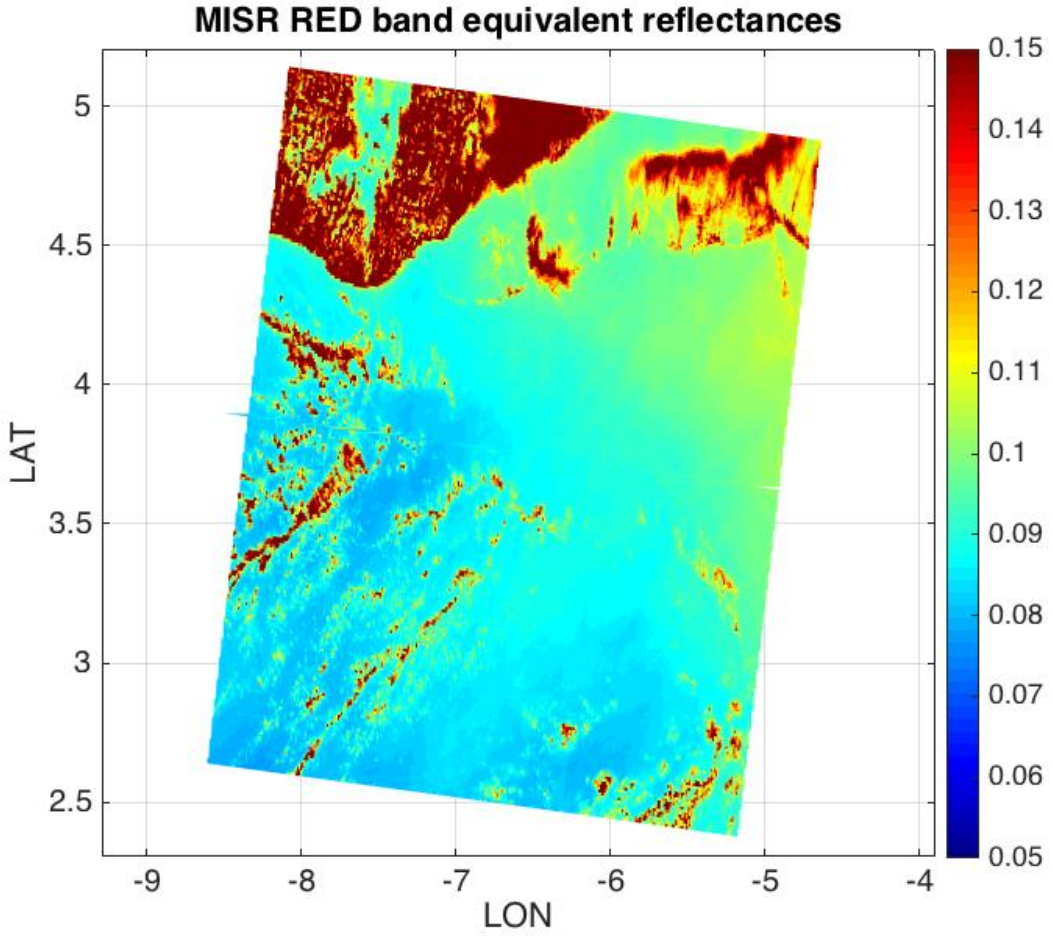

Figure 2 MISR red band equivalent reflectances for the same scene as in Fig. 1. Radiance data does not support the very high-AOD plume indicated by V23 aerosol retrievals.

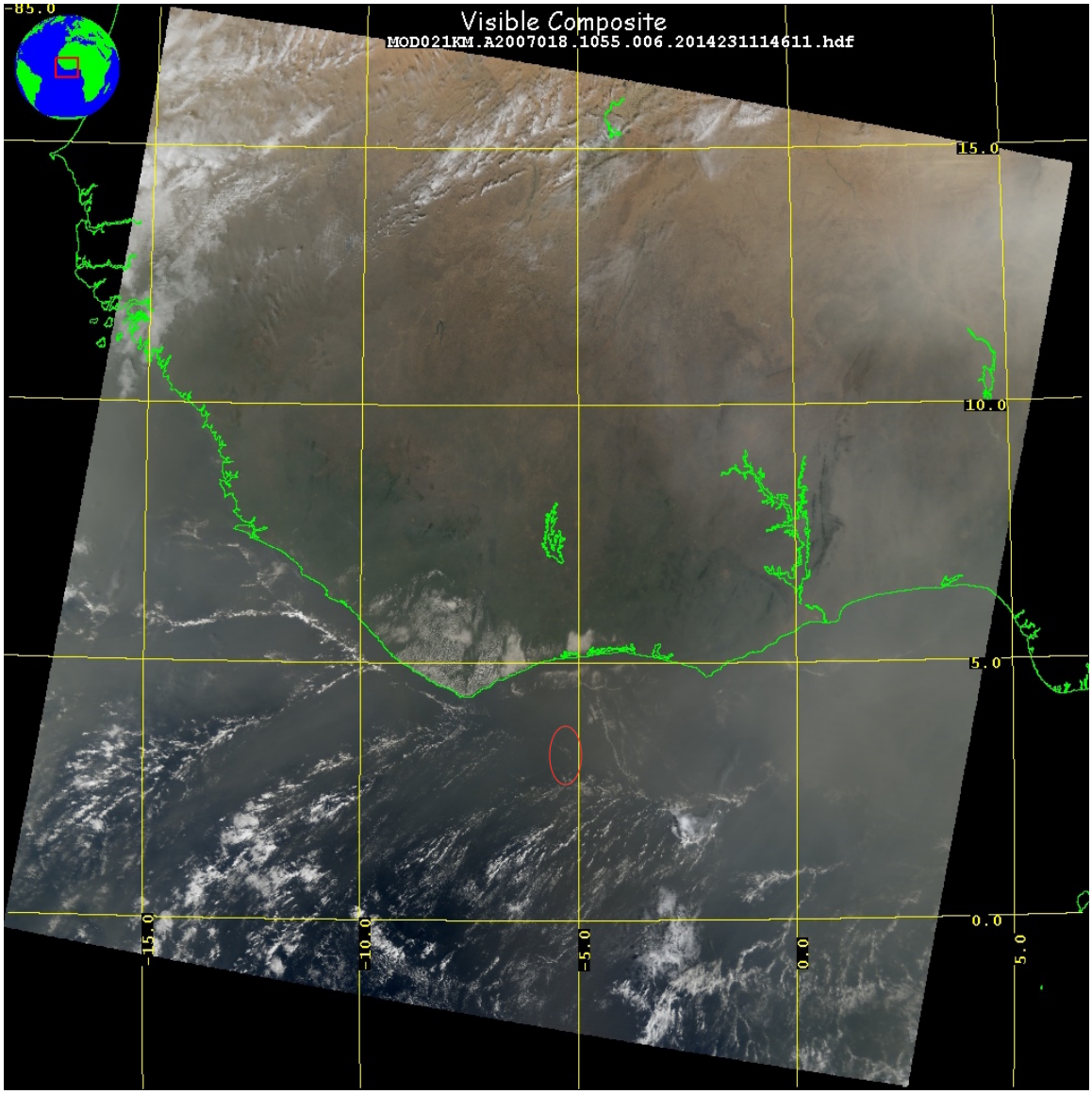

Figure 3 MODIS visible composite for the similar scene as in Fig. 1, with the red oval highlighting an approximate location of the high-AOD and low ARCI retrievals in MISR V23. A substantially thicker aerosol plume is not visible in the MODIS imagery.

*P9L35: Are data in Fig 4 the same data as plotted in Figs 2 and 3?
Re: Yes. We clarified it in the text and in the caption to Fig. 4.
"Another way to look at the difference between the two screening approaches is presented in Fig. 4a, which shows the two-dimensional distribution of average AOD as a function of min(χ^2_{abs}) and ARCI using combined data from January and July of 2007."

"Figure 4 (a) average AOD as a function of ARCI and min(χ^2_{abs}) for the combined months of January and July of 2007…"

*P10L10: These are HUGE differences? Can you compare with anything (e.g. MODIS, AERONET, a model?) to prove this is reasonable? Fig 5 is nice. The "blunders" in the high latitudes (primarily around Antarctica are still glaring.

Re: The difference in global average AOD is large indeed, but the value for the unscreened data is clearly unrealistic. This indicates the impact of ARCI screening on the product. MODIS would give a somehow similar number to the screened V23 product (~0.14).
The speckles of high-AOD in some remote areas are still present but they are addressed by an additional screening procedure not discussed in this paper.

*P10L37: Fig 6. See comment from P2L35: Definitely looks like an a+bxAOD rather than MAX(a,b x AOD).
Re: Yes, this appears to be the case here. We will establish the relationship in the upcoming external validation work.

*P11L12: I am not sure that V23 uncertainties look like V22 uncertainties is useful and or a desired result.
Re: Since AOD uncertainties are reported in the previous V22 MISR aerosol product, it was natural to compare the new V23 uncertainty estimates to the previous ones.

Figures:
Fig. 3: Needs consistent y-axes between pairs of plots
Re: See our response to comment P9L20.

Fig. 7: I am not sure this is a useful figure.
Re: AOD uncertainties have been reported in all versions of the MISR aerosol product. Some readers who have previously looked at this parameter might find it instructive to compare the new predictions with those from V22.