Response to Anonymous Referee #1

Received and published: 21 September 2017

This paper describes a change to the MISR aerosol retrieval algorithm. They select an ensemble of aerosol types and, for each, compute the radiances that would be observed at a range of aerosol optical depths (AOD). Previously, ensemble members were evaluated separately so each gave an AOD and cost, which were then filtered and averaged to calculate the final product. This paper proposes minimising a single cost function (being the sum of the individual cost functions) to find the AOD and it's uncertainty. The technique is rationalised based on two months of observations and is shown to produce more believable uncertainties, on average, than the previous algorithm.

I recommend this paper for publication after minor revisions. The technique proposed is definitely a step in the right direction and the paper is superbly drafted. However, the technique and description thereof could be improved by a more statistical approach. The paper justifies itself with qualitative descriptions of global averages and internal metrics rather than any validation activity, which is common but always disappointing. Specific comments on the paper are listed below, with some minor details collected at the end. The notation PxLy refers to line y of page x.

- My experience is in optimisation. One defines a cost function and selects an algorithm to efficiently search the 'surface' of that function for its global minima. The uncertainty is a measure of the 'width' of that minima in multi-dimensional space (i.e. the magnitude by which a variable could be changed without significantly increasing the cost). The cost function is usually the RMS difference between some modelled value and a measurement. If the model is accurate and the measurement suffers only random noise (of known variance), the minimal value of the cost function will sample a χ^2 distribution, from which one can determine the probability that this measurement fit that model.
  To me, this paper essentially proposes that $f(\tau)$ is a probability density function (PDF) for AOD and that it is normally distributed. It follows that the most likely AOD is the τ that maximises f and the uncertainty is the function's width. The proposed ARCI threshold can then be understood as eliminating retrievals that are exceedingly unlikely. Describing the problem with these basic statistical concepts could vastly simplify the paper, avoiding awkward phrasing like P8L5.
  Re: It is a very valuable observation. We added the following clarification below Eq. 4. "The function f can be interpreted as a probability density function (PDF) for AOD. The most likely AOD is the one that maximizes f (Eq. 4), and the retrieval uncertainty is related to the width of the PDF."
  We also modified the somehow awkward phrasing in P8L5 to read: "Large ARCI, on the other hand, means that for some models sufficiently low χ^2_{abs} were obtained, signifying good agreement with the observations."

- Because this is a fairly straightforward statistical problem, there exists a variety of tools to check that (a) f is in fact a good model of the PDF, (b) f is normally distributed, and (c) the selected aerosol models are an unbiased sampling of the complete state space of real-world aerosols. A brief discussion of some of those points could provide a standardised means to evaluate your assumptions and avoid qualitative judgements, such as the function 'closely resembles a Gaussian' (P7L12).
  Re: In the process of designing and testing the new approach, at one point we did fit a normal distribution to our PDF results. We compared most likely AODs retrieved from PDFs against those retrieved from the fitted normal distributions. The results were in excellent agreement. This exercise gave us confidence

that, at least in those cases that we considered, the PDFs closely resembled Gaussian distributions. However, we though this analysis was too technical to be included in the manuscript. As for point (c), we write in the manuscript that the resulting uncertainty is dependent on the LUT considered in the retrieval (P10L30-35) and that the 74 mixtures currently included in MISR retrieval process are not complete (P4L35-38).

- Are you tabulating f as a function of linear or log τ? Figure 1 uses both as an x-axis, which is misleading. It should be logarithmic as AOD is log-normally distributed (which is clear from the asymmetry about τ_{max} in Fig.1(3)). If you're using linear space, you will underestimate the uncertainty and overestimate the mean.
  Re: All equations in the manuscript use linear τ. In Figure 1a we use the logarithmic scale in the x-axis to better visualize the cost functions at very low τ. Because after inverting the cost functions, at low τ the signal becomes very small, the log scale is no longer necessary. We added additional clarification regarding the x-axis scale in the caption. The distribution in Fig. 1c is close to Gaussian. The misleading resemblance to a log-normal distribution comes from the fact that the PDF is truncated on the left side due to the physical constraint (τ>0.0).

- Why is there no validation of the new algorithm? It seems fairly substantial to move from averaging a few aerosol types per pixel to averaging 74. A few comparisons against AERONET or MODIS would be fine for a paper like this. A simple comparison of V22 vs. V23 would be a start, considering you did it for the uncertainty!
  Re: A validation paper is currently under preparation. It was our intention to designate external validation efforts to a separate publication. One reason for this is that, at the time of preparing this manuscript, we only had two months of data available, which is not enough to obtain sufficient number of collocations with ground based observations. Furthermore, we plan to investigate the new AODs and their pixel-level uncertainties in greater detail, which we feel justifies a separate study.
  Our analysis indicates that the new algorithm leads to AODs that are similar, but not identical, to those obtained using thresholds from V22. However, the uncertainty quantification in the new approach is sufficiently different from V22 to justify a comparison figure (Fig. 7).

- In Sec. 3, you implicitly assume that the choice of aerosol type overwhelms any measurement error. Could Fig. 1 be adapted to show the sensitivity of a χ^2 curve to typical measurement error? I'd expect it to move the curve slightly, but much less than the spread between curves.
  Re: The measurement error is embedded in the calculation of χ^2_{abs} (Eq. 2). The absolute radiometric uncertainty σ_{abs} in V22 is set to 5% of the signal itself for each camera and wavelength (P5L16). We feel that showing the sensitivity of χ^2_{abs} to different levels of σ_{abs} would decrease the clarity of the figure.

P10L24 I'm unhappy with this paragraph.

- L27 I think this is trying to distinguish between a validation activity, which you sadly aren't doing, and an uncertainty estimate, which you are. By definition, uncertainty is a parameter describing the range of values that can be reasonably ascribed to the quantity that is being measured. I believe that provides a 'measure of how far the retrieved AOD deviates from the "truth"'. The distinction is that uncertainty is a prediction of that difference while validation is a direct calculation of it.
  Re: What we are trying to distinguish here is the algorithmic retrieval uncertainty on the one hand, and the uncertainty that comes from comparing a retrieved AOD with ground truth on the other hand. In both

cases we are considering pixel-level information, or individual retrievals, rather than a bulk validation metric like the error envelope. Yes, we are predicting an uncertainty in our algorithm, but this prediction might not necessarily represent the real range of values that are being measured. We are trying to be cautious here and not assign undue credit and value to the algorithm's prediction. A validation activity is required to establish the relationship between the reported uncertainties and the ground truth. Because we think this is a challenging task, we left it for a separate investigation. Our initial results, however, show very promising linkage between our reported uncertainty and the standard deviation of a normally distributed error function.

- It's good to be clear that the estimated uncertainty is sensitive to the way you solve the problem. However, you don't tell the user what to do with that information. I think a rational response at the moment is to avoid MISR data as it's more sensitive to your assumptions than the environment. I can think of three approaches to remedy this:
  1. Give up and declare that your uncertainty values are uncalibrated, providing a pixel-by-pixel assessment of the relative reliability. (I'd recommend that you normalise the values to clarify that their magnitude is not inherently meaningful.)
  2. Show that, despite the algorithm's theoretical sensitivity to your assumptions, the uncertainties you produce are an approximation of the true error. This would be done through a validation activity (e.g. the distribution of $(\tau_{MISR} - \tau_{AERONET})^2 = \sigma^2_{MISR}$ is approximately normal).
  3. Demonstrate that the sensitivity to your assumptions is small. The precise choice of types is a matter for another paper, but it's important to quantify the uncertainty's sensitivity to it. A straightforward way to do so would be re-running the retrieval with a few types removed at random.

  Re: Indeed, at the moment our uncertainty values are uncalibrated. But this is a temporary position that will be resolved in a separate investigation. In order to validate our uncertainties, large comparison statistics against ground truth are required. As mentioned above, at the time of writing we did not have enough data (two months of retrievals) and enough collocations against AERONET to perform a detailed evaluation of the retrieved parameters. This activity will be performed along with the reprocessing of the MISR mission with the new V23 version of the aerosol product.

- Sec. 4 argues that this method is good because it excludes high AOD retrievals. Could you provide some evidence that, for the two months of data you've considered, there were no large aerosol events?
  Re: Figure 5 shows the global distribution of AOD with ARCI screening for January and July of 2007. There are high AOD regions visible off the west coast of Africa and off the coasts of India and China. These are associated with high-AOD events such as dust outflow from Africa, biomass burning, and anthropogenic emissions. We write in the manuscript: (P10L7) "At the same time, climatologically large AODs off the coasts of Africa and South and East Asia are retained, indicating that the new screening method does not unintentionally remove all high AODs that are likely valid."

P3L17 The spread of the MISR ensemble is providing a quantitative insight into the uncertainty in each retrieval due to the assumptions made. While the description of ensemble techniques at L9 is technically correct, ensemble techniques are used to estimate uncertainties that can't be accurately or efficiently calculated by other means. It's exceedingly rare to perturb more than one of the input data, auxiliary parameters, and underlying assumptions. Numerical weather prediction perturbs its input data in order to estimate the sensitivity of a chaotic system. Climate models perturb the auxiliary parameters because they are unknown. MISR perturbs the assumed aerosol type because the radiances available don't fully constrain the problem. MISR doesn't need to perturb the input data as the physics of remote sensing are sufficiently

linear that error propagation does a reasonable job of estimating the uncertainty due to measurement error. Hence, I wouldn't agree that extending ensembles to 'all possible sources of error' would be overly useful. Ensemble techniques are used to quantify uncertainties due to poorly understood, poorly constrained, or exceedingly non-linear error sources.

Re: We modified this sentence to read: "Such an approach, if extended to all poorly quantifiable nonlinear sources of error and physically plausible realizations of parameter space, has the potential of providing a robust and comprehensive measure of retrieval uncertainty in the manner suggested by Povey and Grainger (2015)."

P8L38 Within this paper, the only evidence that the cloud filtering is effective is showing that mean AOD is lower. MISR is on the same platform as a MODIS, so you have the ability to check if your cloud flagging spatially agrees with them. That would be rather more convincing than the distribution of a month's observations presented in Fig. 5.

Re: Yes, it could potentially be convincing to compare our screening method against MODIS. However, comparing different cloud screening techniques between satellite instruments, even on the same platform, is quite challenging and in our opinion it would extend beyond the scope of this study. MISR and MODIS have different spectral bands with different calibrations, different spatial resolutions, and the data are projected differently. The Global Energy and Water cycle Experiment (GEWEX) has an extensive report that describes such instrumental differences and compares their cloud products available online (http://climserv.ipsl.polytechnique.fr/gewexca/). The point being made in the manuscript is that the ARCI-based retrieval screening provides first line of defense against cloud-contaminated retrievals. Additional screening steps using other types of information are applied to filter out more retrievals potentially contaminated by clouds. These cloud-screening steps will be described in a separate publication.

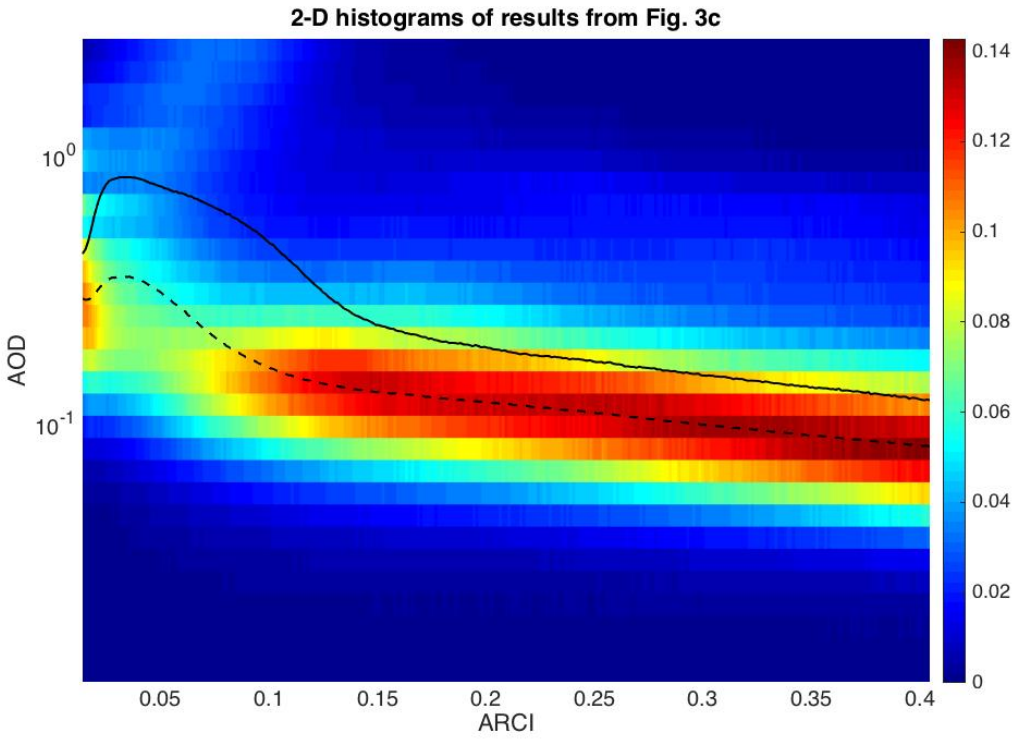
Fig.3 (b) is rather concerning. Do the peaks in retrieval count correspond to the divisions of your LUT? Also, could 3(a) and (c) be shown as 2-D histograms with the mean overplotted? Your argument would be stronger if the decrease in mean AOD with increasing ARCI is due to a decreased prevalence of large AOD (the cloud-contaminated retrievals) while the variation with χ^2 is more uniform.

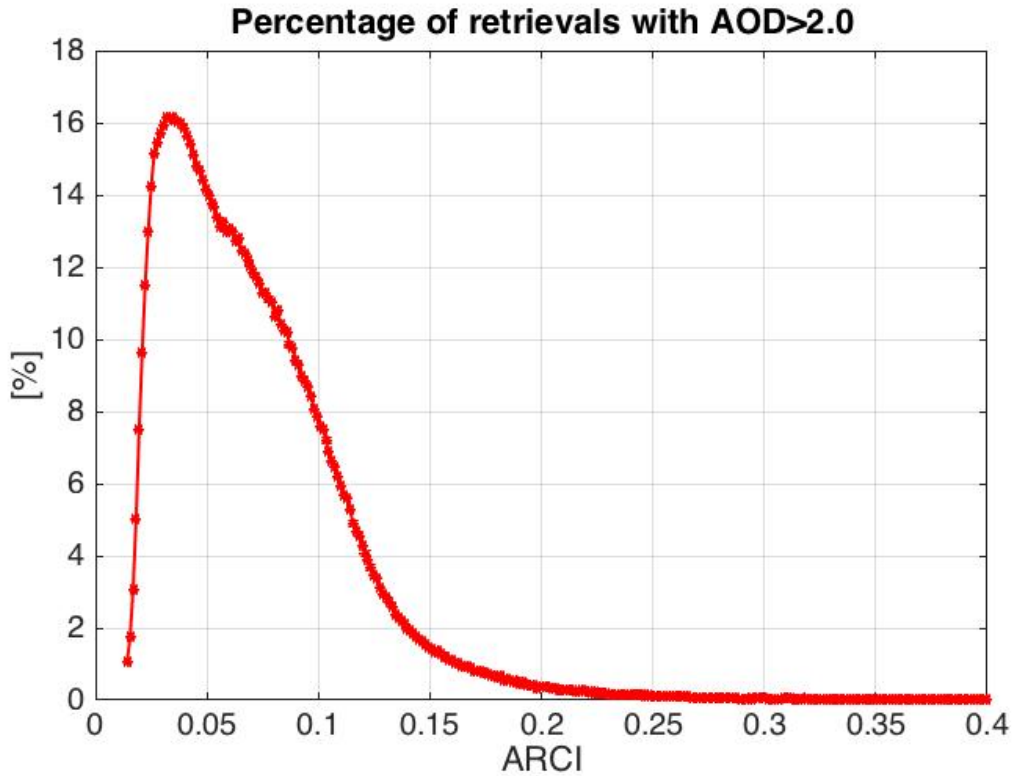
Re: We do see certain clustering around specific min(χ^2) values in our dataset, which gives rise to the small wiggles seen in Fig. 3b. This is probably related to the finite AOD gridding of our LUT, which is 0.025 throughout most of the AOD range. We plan to investigate this feature in greater detail in the future. Furthermore, the wiggles in Fig. 3b become apparent only because of very fine sampling of the min(χ^2) space. Our interval is 0.025, which results in 200 data points for min(χ^2) between 0 and 5.

We created a 2-D figure with results from Fig. 3c by plotting normalized histograms of AOD at each ARCI level. An example is presented below in Figure 1. The black solid and dashed lines are the mean and the median AODs. The figure does show decreasing number of high AODs with increasing ARCI, but the results are not as clearly visible as in Fig. 3 in the manuscript. Another useful metric showing that the number of cloud-contaminated high-AOD retrievals is decreasing with increasing ARCI is the percentage of retrievals with AODs higher than 2.0. This is presented below in Figure 2. The percentage of high-AOD retrievals decreases from the top 16% at ARCI=0.03 to about 1% at ARCI=0.15. Figure 2 below is simple and clearly conveys the message, but we decided a description in the text was sufficient to strengthen our argument.

"In the first regime, the average AOD is highly sensitive to the specific value of ARCI, characterized by a sharp decrease in AOD with increasing ARCI between about 0.03 and 0.13. This suggests that a decreasing number of cloud-contaminated, high-AOD retrievals are included in the average as the

ARCI is increased. Indeed, the percentage of retrievals with AOD higher than 2.0 reaches its peak, 16%, at ARCI equal to 0.03, and decreases to about 2% when ARCI is 0.13. In the second regime…"





P9L2 This paragraph ascribes the variations in Fig. 3 at low χ^2 or ARCI to poor sampling. That implies that there should be retrievals there but you didn't see them. Very low χ^2 implies a very close fit to observations, which is unlikely, and very low ARCI implies a very unlikely fit, which should happen infrequently if the ensemble of aerosol types was well-chosen. Hence, I'd ascribe the sharp variations in Fig. 3 in those regions to scenes that are poorly suited to this retrieval.

Re: We agree with this statement. We ascribed the variations at low χ^2 or ARCI to poor sampling without providing an explanation of why the sampling in these regimes is low. We did not want to put too much emphasis in our analysis to these low χ^2 or ARCI regimes, as they are not very relevant to our main arguments. We did, however, change the phrasing in this paragraph:

"After a rapid initial drop related to a similar rapid increase in sampling…"

"After excluding the initial fluctuation for extremely small ARCI related to poor sampling…"

P9L19 I wouldn't say that the trend in AOD is statistically robust. I'd say that the shape of 3(c) isn't evident in 3(d), so we don't ascribe the kink in the former to a change in frequency.

Re: Agreed. We changed this sentence to read: "The retrieval count decreases slowly with increasing ARCI (Fig. 3d), indicating that the observed trends in the average AOD cannot be ascribed to a change in frequency."

Fig.4 This is a superb figure and deserves more attention than Fig. 3. However, the caption is unclear if it is plotting the same data as in Fig. 3.

Re: We clarified the data used in the caption and in the text:

"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…"

P10L12 Any idea why cloud contamination is a function of latitude? Does the ARCI threshold need to vary with latitude?

Re: Global cloud fraction has a strong latitudinal component due to the patterns of global atmospheric circulation. We have not investigated possible variations of the ARCI threshold with latitude. We will consider this possibility in the future.

P12L19 You didn't provide a 'strong statistical foundation'. You justified the ARCI threshold by the shape of the distribution of AOD. Statistics would calculate a theoretically sensible value of ARCI based on typical noise and a very large ensemble of aerosol types.

Re: Agreed. We changed this sentence to read:

"Although this screening method does not eliminate all AOD outliers, it is superior to the previously used thresholds in V22 of the MISR aerosol product."

- Finally, I would prefer it if the paper and any data files released clearly describe the retrieved product as 'ensemble mean AOD'. Evaluating a range of aerosol types is an excellent way to sample the unconstrained parts of state space (such as refractive index). Providing an ensemble of results to the user illustrates what the data constrains and what it doesn't. However, a combination of ensemble members doesn't necessarily have a physical meaning. To use an example from a related problem, a thick but high cloud can produce the same TOA thermal radiance as a thin and low one. Giving the user both results shows that both are possible. An ensemble mean, though, gives a medium-thickness layer midway through the atmosphere, which is inconsistent with the data.

  Re: The new V23 data product labels the retrieved AOD as, simply, "Aerosol_Optical_Depth". To a sophisticated user, the idea that this is essentially a "ensemble mean AOD" is a useful concept, which is one of the reasons for writing this manuscript. However, as this AOD is the one the MISR project would like the majority of users to work with, we elected to eliminate the jargon and provide a simpler designation for this field.

A few more minor points:

P4L6 Perhaps 'The previous MISR dark water algorithm' would be a more informative title to someone skimming the paper?

Re: We changed the title of the second section to: "Previous MISR V22 dark water algorithm"

P5L2 reflectance is defines defined as

Re: Corrected.

P5L26 Considering you don't define them, and their precise definition is unimportant to this paper, perhaps remove specific references to the now neglected χ^2, parameters?

Re: We want to make sure that the readers are aware of additional metrics and thresholds used in V22 processing. This is important as the new approach simplifies the process considerably and makes it more transparent.

P6L34 'turns out to be' is rather colloquial. Perhaps 'and will be shown to produce superior results to the original algorithm'?

Re: Agreed. We modified this sentence to read: "Furthermore, it results in a single parameter that enables screening of retrieval blunders and AOD outliers and which outperforms results derived using the original V22 thresholds."

P7L4 If these are continuous functions of τ, you are presumably interpolating as the LUT is discrete. What are you interpolating — ρ; χ^2; or f?

Re: We interpolate χ^2

Response to 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.

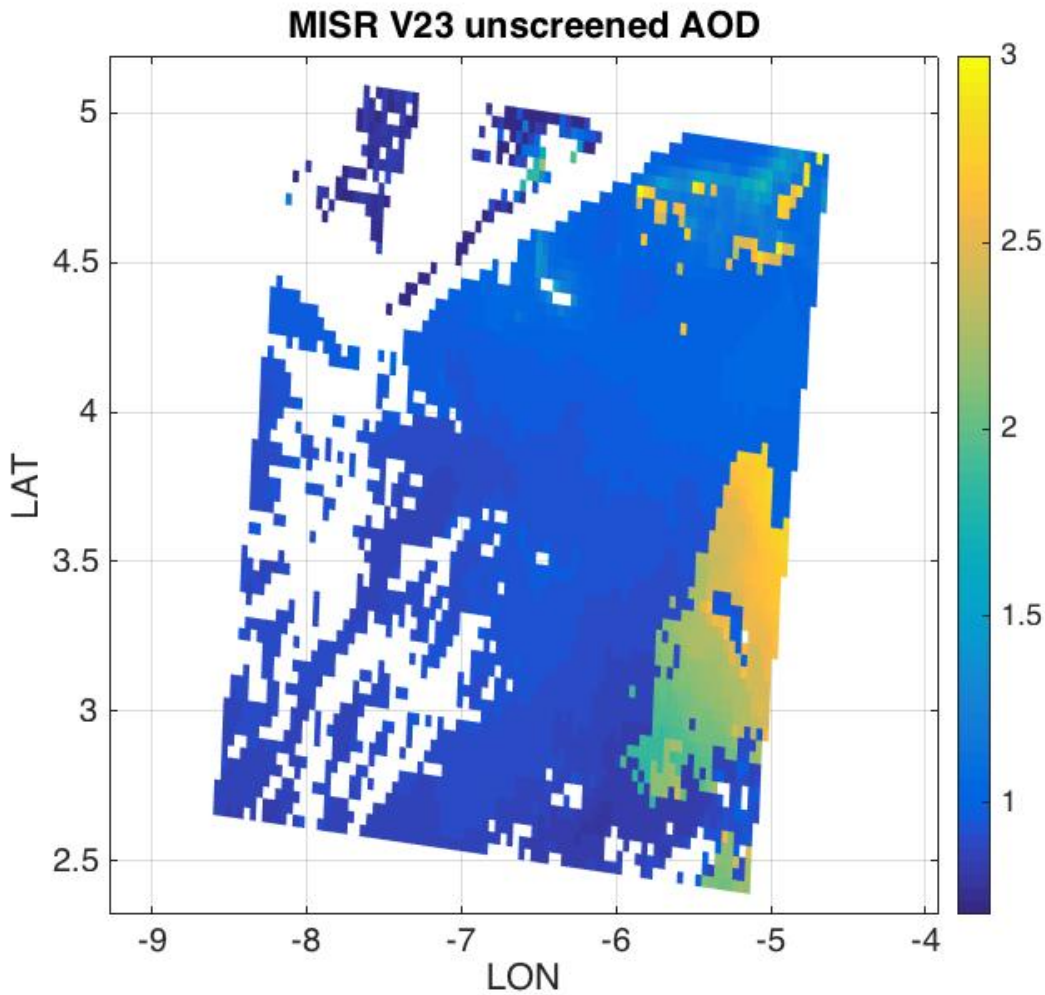


Figure 1 MISR V23 unscreened AOD from orbit 37689, blocks 87-88, time: January 18, 2007, 10:58 UTC. The high-AOD retrievals in the center right of the image have low ARCI and are therefore screened out in the final product. These retrievals, however, have min(χ^2_{abs}) values below 2.0 and therefore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

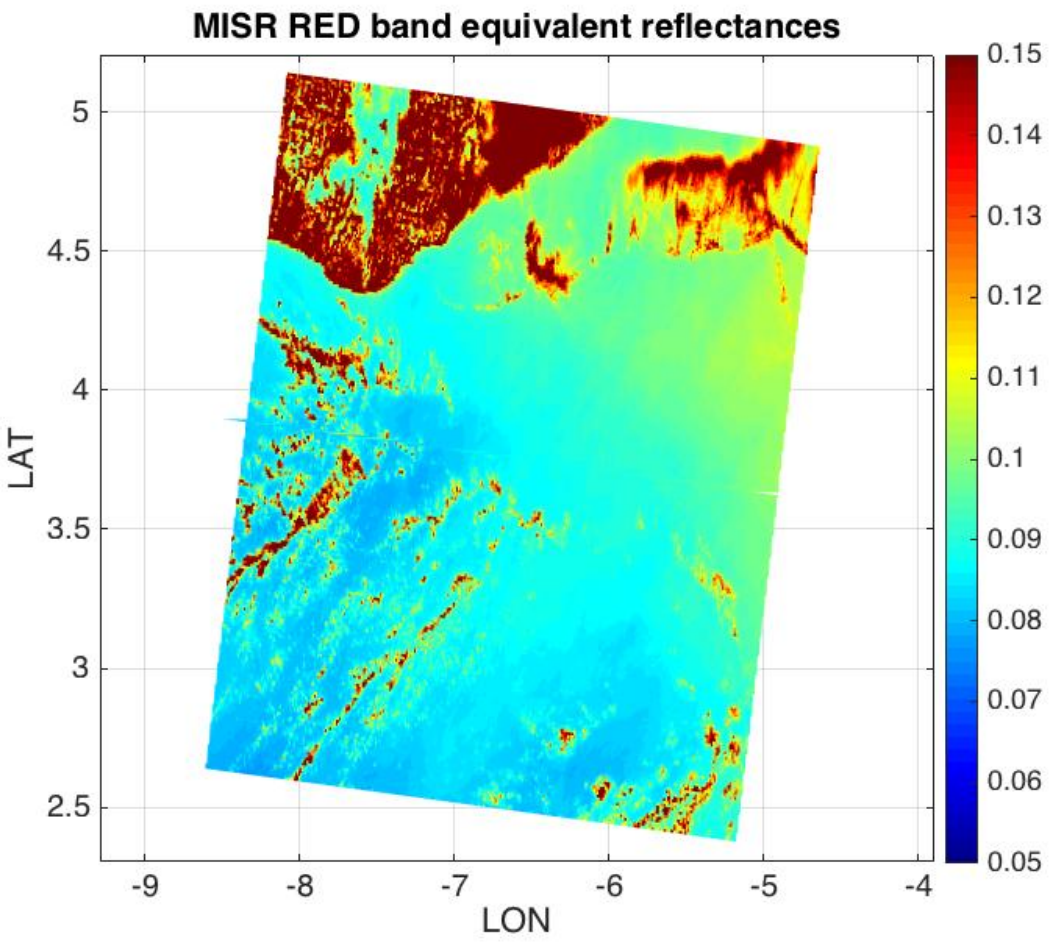


Figure 2 MISR red band equivalent reflectances for the same scene as in Fig. 1. Radiance data does not support the very high-AOD plume indicated by V23 aerosol retrievals.

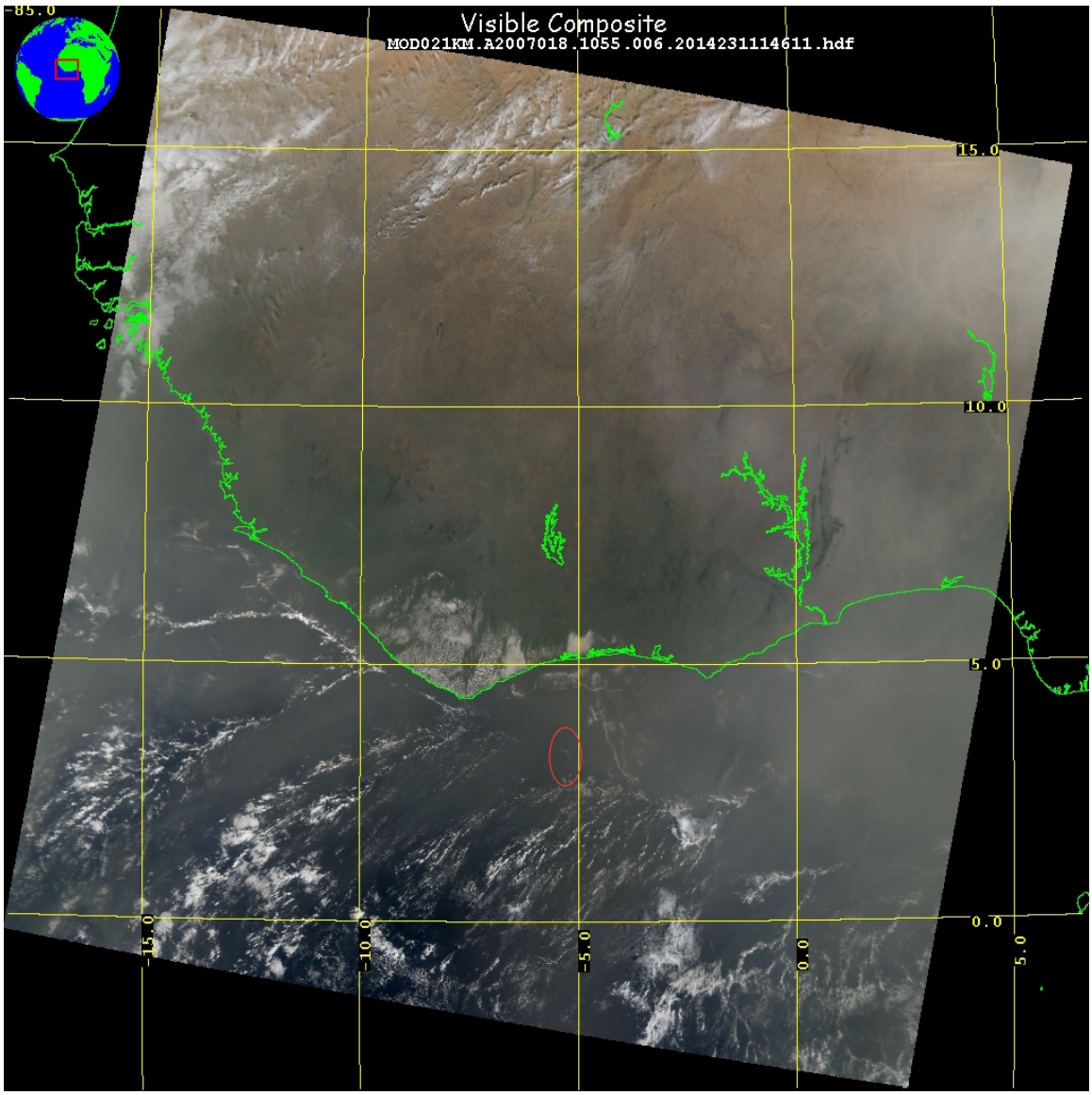


Figure 3 MODIS visible composite for the similar scene as in Fig. 1, with the red oval highlighting an approximate location of the high-AOD and low ARCI retrievals in MISR V23. A substantially thicker aerosol plume is not visible in the MODIS imagery.

*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.

Response to Anonymous Referee #3

Received and published: 21 September 2017

This manuscript details a new method to retrieve AOD and pixel-level AOD uncertainty from the Multi-angle Imaging SpectroRadiometer. The manuscript is well-written, and it is likely that this new method will serve the MISR aerosol community well in the future. However, I can not support publication of this work without at least a cursory attempt at validation, which this paper is sorely missing.

Major Comment: Although I believe that this new method probably represents a substantial improvement for aerosol remote sensing from MISR, it is incumbent on the authors to prove this. The authors claim that not all sources of error are included, therefore no comparison of the AOD differences and pixel-level uncertainties (against MAN or AERONET, for instance) should be done. If this new method is going to be implemented in the next version of the MISR aerosol retrieval algorithm (and going to be published here), it should be validated first. A comparison of new algorithm results with old algorithm results does not replace real validation.
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 not have been unsatisfactory.

Minor Comment:
The authors mention (5 times) that several of the thresholds from the current version of the MISR aerosol retrieval algorithm are arbitrary. Please refrain from so much hyperbole in the manuscript, as most empirical thresholds could be considered arbitrary (including your own ARCI threshold).
Re: We eliminated most occurrences of the phase "arbitrary thresholds" from the manuscript and substituted it with "empirical thresholds". The one remaining case is on page 12, line 14: "This approach allows for a consistent calculation of AOD and AOD uncertainty without the need for screening acceptable mixture solutions based on a complex interplay of multiple, and somewhat arbitrary, thresholds."

New approach to the retrieval of AOD and its uncertainty from MISR observations over dark water

Marcin L. Witek[1], Michael J. Garay[1], David J. Diner[1], Michael A. Bull[1], and Felix C. Seidel[1]

[1]Jet Propulsion Laboratory, California Institute of Technology, 4800 Oak Grove Drive, Pasadena, CA 91109, USA

Correspondence to: Marcin L. Witek (marcin.l.witek@jpl.nasa.gov)

Abstract

A new method for retrieving aerosol optical depth (AOD) and its uncertainty from Multi-angle Imaging SpectroRadiometer (MISR) observations over dark water is outlined. MISR's aerosol retrieval algorithm calculates cost functions between observed and pre-simulated radiances for a range of AODs (from 0.0 to 3.0) and a prescribed set of aerosol mixtures. The previous Version 22 (V22) operational algorithm considered only the AOD that minimized the cost function for each aerosol mixture, then used a combination of these values to compute the final, "best estimate" AOD and associated uncertainty. The new approach considers the entire range of cost functions associated with each aerosol mixture. The uncertainty of the reported AOD depends on a combination of a) the absolute values of the cost functions for each aerosol mixture, b) the widths of the cost function distributions as a function of AOD, and c) the spread of the cost function distributions among the ensemble of mixtures. A key benefit of the new approach is that, unlike the V22 algorithm, it does not rely on empirical thresholds imposed on the cost function to determine the success or failure of a particular mixture. Furthermore, a new Aerosol Retrieval Confidence Index (ARCI) is established that can be used to screen high-AOD retrieval blunders caused by cloud contamination or other factors. Requiring ARCI≥0.15 as a condition for retrieval success is supported through statistical analysis and outperforms the thresholds used in the V22 algorithm. The described changes to the MISR dark water algorithm will become operational in the new MISR aerosol product (V23), planned for release in 2017.

1. Introduction

Uncertainty estimation in satellite remote sensing is a highly challenging endeavor that requires diligent assessment of many potential sources of error. Some of these errors are random and well understood and generally follow a traditional error propagation pathway. Other errors—which in underconstrained retrievals of geophysical quantities can be the dominant sources of uncertainty—are not readily assessed as they stem from various approximations and assumptions that lead to non-linear systematic responses. Additionally, there are sources of error that cannot be quantified at all that are attributable to the resolution of an instrument and variability at scales smaller than those observed.

Recently, Povey and Grainger (2015) provided a detailed overview and classification of possible sources of error in satellite retrievals. They clearly distinguish between pixel-level uncertainties associated with individual retrievals and bulk uncertainty metrics originating from validation studies. The two terms, however, are often not properly differentiated in the literature. This is in part due to a lack of awareness about the issue as most studies focus on external validation of satellite data products, which often do not contain information about the uncertainties associated with individual retrievals. This status quo started changing due to growing pressure from the data assimilation community, which requires information on pixel-level uncertainties in order to make proper use of satellite-derived geophysical quantities within the framework of a numerical model.

An example of a remotely sensed geophysical parameter that is being increasingly used in data assimilation applications is the aerosol optical depth (AOD)(Lynch et al., 2016; Shi et al., 2011, 2013; Zhang et al., 2008; Zhang and Reid, 2006, 2009, 2010). AOD represents the total extinction by aerosol particles in the atmospheric column from the surface to the top of the atmosphere. Model data assimilation generally requires a value of AOD and an uncertainty associated with this value in order to determine whether or by how much the model should be adjusted to agree with the retrieval. AODs are readily available in many satellite data products, such as those from the Moderate Resolution Imaging Spectroradiometer (MODIS), the Multi-angle Imaging SpectroRadiometer (MISR), and the Visible Infrared Imaging Radiometer Suite (VIIRS), whereas individual retrieval uncertainties, in many cases, are not. Pixel-level information often consists of retrieval quality assurance flags, which categorize the "usability" of a retrieval based on a qualitative judgment of the algorithm performance. For example, MODIS Collection 6 uses four coarse labels to communicate retrieval quality, ranging from 0 to 3, with 0 being the least trusted and 3 indicating the highest quality (Hubanks, 2015; Levy et al., 2013; Remer et al., 2005). Other metrics conveying some useful information about AOD retrieval uncertainty are the proximity of clouds and cloud coverage in the retrieval region (Shi et al., 2011, 2013, 2014; Witek et al., 2013; Zhang and Reid, 2006). 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.

The most frequently used metric that quantitatively characterizes the quality of a particular AOD dataset as a whole is the error envelope (EE)(Bréon et al., 2011; Garay et al., 2016; Kahn et al., 2010; Levy et al., 2010, 2013; Limbacher and Kahn, 2014; Omar et al., 2013; Remer et al., 2008; Sayer et al., 2012, 2013; Witek et al., 2013). EE results from a validation study where a particular satellite AOD dataset is compared against another AOD dataset, typically ground-based information from the Aerosol Robotic Network (AERONET) (Dubovik et al., 2000; Holben et al., 1998), which is considered to represent the "truth". Taking the general form of $\pm(a+b\times AOD)$ (or $\max[\pm a, \pm(b\times AOD)]$), where a and b are empirically determined constants, EE represents the range required so that about 68% (one standard deviation) of the matched data agree, providing an overall characterization of the entire dataset. The EE is often regarded as an expected retrieval error, or confidence envelope. External validation is fundamental to assessing a dataset's overall quality and serves as a useful guide for identifying the presence of systematic errors in the dataset

(e.g., Kahn et al., 2010). The EE, which describes the performance of the dataset as a whole relative to a reference dataset, however, does not represent and should not be confused with the uncertainty characteristics of an individual retrieval.

More appropriate attempts to address the true AOD retrieval uncertainty involved sensitivity tests of the retrieval algorithm with respect to varying external factors such as lower boundary conditions or aerosol microphysical properties (Kahn et al., 2001; Kalashnikova and Kahn, 2006; Sayer et al., 2016). These studies were limited in scope as they were unable to address all possible sources of error. Povey and Grainger (2015) suggested the use of ensemble techniques as a comprehensive means of quantifying uncertainties in satellite remote sensing of the environment. Each member of the ensemble adds a random perturbation to the measurements, ancillary parameters, and underlying retrieval assumptions in order to comprehensively map the probability distribution of the retrieved quantity. Even though such an approach would face significant conceptualization and computational challenges, especially in operational data processing, there are already examples of ensembles being used in aerosol remote sensing. One example is MISR (Diner et al., 1998), which employs many different particle microphysical mixtures as part of its operational aerosol retrieval process (Kahn et al., 2001; Martonchik et al., 1998). The resulting spread of AOD solutions for this ensemble of aerosol mixtures offers quantitative insight into the uncertainty of the individual retrieval. Such an approach, if extended to all poorly quantifiable nonlinear sources of error and physically plausible realizations of parameter space, has the potential of providing a robust and comprehensive measure of retrieval uncertainty in the manner suggested by Povey and Grainger (2015).

This study describes a new approach to determining AODs and AOD uncertainties in MISR retrievals. MISR is an instrument aboard the National Aeronautics and Space Administrations (NASA) Earth Observing System (EOS) Terra satellite that performs radiometric observations of the Earth using nine pushbroom cameras pointing at nine different angles (Diner et al., 1998, 2005). All nine cameras observe the same area on Earth within seven minutes, at four different wavelengths. The multi-angle viewing capability allows MISR to sample portions of the scattering phase function simultaneously, providing information that helps distinguish between different aerosol types (Kahn et al., 2001; Kalashnikova and Kahn, 2006). The retrieval process is based on the minimization of a cost function evaluated between the instrument observations and pre-calculated radiometric look-up tables (LUTs) that are in turn based upon a prescribed set of aerosol mixtures. Version 22 (V22) of the algorithm, which has been in operational production since December 2007, considers 74 aerosol mixtures with different microphysical properties that are intended to represent typical atmospheric conditions found on Earth (Kahn et al., 2009, 2010). MISR has two aerosol processing pathways, one for dark water (oceans, seas, deep lakes) and the other for the land surface (Martonchik et al., 2002). The modifications described in this study currently apply only to the dark water algorithm (Kalashnikova et al., 2013; Witek et al., 2013).

The paper is organized as follows. Section 2 describes the V22 operational approach to determining AODs and their uncertainties in MISR aerosol retrievals over dark water. In section 3 critical modifications and changes, which form the basis of a new methodology, are presented.

Section 4 introduces an important new metric that is employed to assess the quality of retrievals. This criterion is used for distinguishing between "good" and "poor" retrievals. Section 5 offers statistical analysis of the new AOD retrieval uncertainty and comparisons against the V22 approach. Finally, a short summary of the study follows in section 6.

2. Previous MISR V22 dark water algorithm

A detailed description of the MISR retrieval strategy is described in the Level 2 Aerosol Retrieval Algorithm Theoretical Basis document (Diner et al., 2008). The MISR aerosol retrieval algorithm follows two separate lines of processing depending on the surface type: dark water and land. The two retrieval types are independent of each other and largely rely on different physical and empirical assumptions. Only the dark water algorithm (Kalashnikova et al., 2013) is considered in this study. Here some key elements of the V22 algorithm relevant to the new approach are reviewed.

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. One-dimensional radiative transfer (RT) theory is sufficient for determining the relationship between top-of-atmosphere (TOA) radiances and AOD. However, this calculation assumes an aerosol model that specifies the particle size distribution, shape, complex refractive index, and vertical distribution. Additional assumptions are made about the gaseous concentration in the atmosphere (ozone absorption, Rayleigh scattering) and surface whitecap fraction (i.e., the area of the surface covered by white foam from breaking waves). MISR's ability to observe multi-angle radiances, which are in large part governed by the shape of the aerosol scattering phase functions, provides a wealth of information with which to refine aerosol retrievals over dark water (Kalashnikova and Kahn, 2006).

The MISR aerosol retrieval algorithm relies on a LUT generated for a predefined set of mixtures with known optical properties. The V22 operational algorithm defines 74 aerosol mixtures, each of which consists of up to three unique particle types having prescribed optical and microphysical properties (Kahn et al., 2010; Kahn and Gaitley, 2015). The 74 mixtures consist of combinations of eight primary particle types. The MISR particle types and mixtures are designed to represent several compositional categories typically found in the atmosphere, such as sea spray, sulfate/nitrate, mineral dust, carbonaceous, and urban soot aerosols. Recently, Kahn and Gaitley (2015) provided a thorough verification of MISR's aerosol type retrieval capabilities and Lee et al. (2016) demonstrated that the MISR aerosol particle climatologies regionally showed good agreement with the results of chemical transport models. The current 74 mixture set, however, is not complete and has some documented deficiencies (Kahn et al., 2010; Kalashnikova et al., 2013; Limbacher and Kahn, 2014). A more comprehensive set of aerosol types and mixtures will be considered in a future release of the MISR's aerosol algorithm.

For each of the 74 mixtures, forward RT calculations are performed to provide top-of-atmosphere (TOA) radiances for the 36 MISR channels (9 angles and 4 wavelengths). These

radiances are stored in the form of "equivalent reflectances" in the Simulated MISR Ancillary Radiative Transfer (SMART) Dataset, where equivalent reflectance is defined as

$$\rho = \frac{\pi L}{E_0}, \tag{1}$$

where L is the radiance and E_0 is the exo-atmospheric solar irradiance determined for each MISR spectral band. The RT calculations of TOA radiances are carried out for discrete values of mixture optical depth, from 0 to 3, referenced to MISR's 558 nm (green) band for all plausible combinations of view and solar geometries. The simulations incorporate a modified linear mixing theory for mixtures containing more than one aerosol type, a wind-speed-driven glitter and whitecap model, ozone correction, and Rayleigh scattering (Abdou et al., 1997). The modeled TOA radiances are then directly compared with the MISR observations. The criterion used to find the best-fitting optical depth for a particular mixture is minimization of the reduced χ^2_{abs} parameter, calculated as a function of green-band AOD (τ)

$$\chi^2_{abs}(\tau) = \frac{\sum_{l=1}^{4} w_l \cdot \left[\sum_{j=1}^{9} v(l,j) \cdot \frac{\left[\rho_{MISR}(l,j) - \rho_m(l,j)\right]^2}{\sigma^2_{abs}(l,j)} \right]}{\sum_{l=1}^{4} w_l \cdot \left[\sum_{j=1}^{9} v(l,j) \right]}. \tag{2}$$

In Eq. 2, ρ_{MISR} are MISR equivalent reflectances, ρ_m are modeled TOA equivalent reflectances for a given aerosol mixture, and σ_{abs} are the absolute radiometric uncertainties in ρ_{MISR} calculated as $\sigma_{abs}(l,j) = 0.05 \max\left(\rho_{MISR}(l,j), 0.04\right)$. The summation index l is over the 4 MISR wavelengths and j is over the 9 MISR cameras. The parameter $v(l,j) = 1$ if a valid value of $\rho_{MISR}(l,j)$ exists, and is set to 0 otherwise. The weights w_l are always equal 1 for the red and NIR bands, and are ≤ 1 for the blue and green bands, depending on τ. These weights allow individual bands to contribute varying amounts to the χ^2_{abs} calculation as a function of optical depth. For example, at low τ (<0.5) the MISR 446 nm (blue) and 558 nm (green) bands are not used in χ^2_{abs} calculations since at these wavelengths and τ values the water leaving radiance could be a significant contributor to the TOA signals.

For each mixture, the best fitting value of τ_{mix} is taken to be the value that minimizes χ^2_{abs} using a parabolic fitting approach (Diner et al., 2008). Additional parameters are used to determine the goodness of fit of the particular aerosol mixture to the MISR data. Those are χ^2_{geom}, χ^2_{spec}, and $\chi^2_{\max dev}$ which are calculated at the previously obtained value of τ_{mix}. Definitions of these parameters can be found in Diner et al. (2008) and Kalashnikova et al. (2013). An aerosol mixture is considered "successful" if all four metrics χ^2_{abs}, χ^2_{geom}, χ^2_{spec}, and $\chi^2_{\max dev}$ do not exceed certain empirically established thresholds. In V22, these thresholds are set to 2, 3, 3, and 5, respectively.

As a result of this procedure, each retrieval region is assigned a set of mixtures and associated AODs that pass all the threshold criteria. In the special case when none of the 74 mixtures pass the threshold tests, the retrieval is considered unsuccessful and no AOD is reported. In most cases, however, multiple mixtures are deemed successful. They typically have somewhat different τ_{mix} values corresponding to their minimum χ^2_{abs}. The arithmetic mean of all AODs from the passing mixtures is reported as the "best estimate" AOD in the V22 Level 2 aerosol product with the field name "RegBestEstimateSpectralOptDepth". The retrieval uncertainty, with the field name "RegBestEstimateSpectralOptDepthUnc", is determined from the standard deviation of the AODs from the passing mixtures. In the case where only a single mixture is successful, the uncertainty is determined directly from the parabolic fit for that mixture (Diner et al., 2008).

A critical aspect of the retrieval process and its outcome is its dependence on a number of empirically determined thresholds. The specific numerical values for the χ^2_{abs}, χ^2_{geom}, χ^2_{spec}, and $\chi^2_{\max dev}$ thresholds were chosen based on pre-production tests aimed at eliminating obvious blunders while maintaining adequate spatial coverage. However, the reported AOD and its uncertainty are directly linked to these thresholds, which entails a certain degree of subjectivity. For example, a mixture and AOD combination resulting in a χ^2_{abs} value of 1.99 would be considered successful, while a mixture/AOD combination with a χ^2_{abs} value of 2.01 would not be. Alternatively, mixtures with very different AODs, for example a non-absorbing and an absorbing mixture, might both be considered successful and be included in the uncertainty calculation, but have dramatically different χ^2_{abs} values, something which is not taken into account when determining the uncertainty. Mitigating such issues was an important driver for developing a new approach to AOD determination and its uncertainty for MISR dark water retrievals in a more objective manner.

3. New approach to AOD retrieval and its uncertainty

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. A more desirable approach would minimize the reliance on empirical thresholds. In this section, a new method is described that meets this objective and simplifies the retrieval of the "best estimate" AOD and its associated uncertainty. Furthermore, it results in a single parameter that enables screening of retrieval blunders and AOD outliers and which outperforms results derived using the original V22 thresholds. This algorithm revision has been implemented in the software used to generate the next version of the MISR operational aerosol product, V23, scheduled for public release in 2017.

The new method relies solely on the χ^2_{abs} values calculated using Eq. 2. The other goodness-of-fit metrics are retained solely for diagnostic purposes. Extensive testing showed that AOD

selection in V22 was governed primarily by χ^2_{abs}, with the other parameters typically having little effect due to the magnitude of their thresholds, except in a limited set of cases. The key elements of the new method are visualized in Figure 1 using actual MISR data from a randomly selected case. First, the values of χ^2_{abs} for each mixture are calculated as continuous functions of τ. The result is then inverted, yielding the distribution of $1/\chi^2_{abs}$. Taking the reciprocal has two primary benefits. First, it gives a smaller weight to retrievals with large values of χ^2_{abs} that represent poor agreement between the model and the MISR observations. Second, the distribution of $1/\chi^2_{abs}$ tends to look Gaussian, with a peak at τ_{mix}. In the next step, these functions are averaged over all *N*=74 mixtures, leading to:

$$f(\tau) = \frac{1}{N}\sum_{m=1}^{N}\frac{1}{\chi^2_{abs,m}(\tau)}. \quad (3)$$

The position of the peak of the average distribution *f* is the new retrieved AOD

$$AOD = \tau\left(\max\left(f(\tau)\right)\right). \quad (4)$$

The function *f* can be interpreted as a probability density function (PDF) for AOD. The most likely AOD is the one that maximizes *f* (Eq. 4), and the retrieval uncertainty is related to the width of the PDF. The function *f*, which in most cases closely resembles a Gaussian (normal) distribution, has a peak that is narrow or wide depending on how closely the individual τ_{mix} from the 74 mixtures are clustered. Furthermore, because the absolute values of $1/\chi^2_{abs,m}$ from each mixture contribute to the overall shape of the PDF, aerosol models fitting the observations well (low χ^2_{abs}) dominate the shape of *f* and the position of its peak, whereas mixtures with poor fits (high χ^2_{abs}) contribute less. The retrieval uncertainty (σ) is determined from the full width at half maximum (FWHM) of *f*, and scaled to a standard deviation under the assumption that the functional form of *f* in the vicinity of its peak can be approximated by a Gaussian distribution:

$$\sigma = \frac{FWHM(f)}{2\sqrt{2\ln 2}} \approx \frac{FWHM(f)}{2.3548}. \quad (5)$$

Equations 4 and 5 form the backbone of the new approach for determining AODs and their uncertainties in MISR retrievals over dark water in the V23 algorithm. One important benefit of the method is that it does not rely on empirically determined thresholds. In all cases, all 74 mixtures contribute to the retrieved AOD, but the amount they contribute depends on how well they agree with the MISR observations. The retrieval uncertainty is then related to the degree to which the AODs associated with the entire ensemble of aerosol mixtures cluster around a specific AOD. If all mixtures are consistent with the same AOD and are highly sensitive to its specific value, the peak in *f* will be narrow and the uncertainty low. If mixtures disagree as to a single value of AOD, or the χ^2_{abs} parameter is relatively insensitive to the AOD, the distribution will be broad and the reported uncertainty will be larger.

While the width of the average distribution $f(\tau)$ contains information about the retrieval uncertainty, the peak of the distribution, $\max(f(\tau))$, has an additional important benefit: it can be utilized as a retrieval screening parameter. $\max(f(\tau))$ represents the overall agreement of the TOA equivalent reflectances from the aerosol mixtures in the LUT with the MISR observations, which can be considered a measure of the confidence in the retrieval. $\max(f(\tau))$ is designated the "Aerosol Retrieval Confidence Index", or ARCI. Low ARCI implies that generally high χ^2_{abs} were obtained, indicating that the aerosol models fit the MISR observations poorly. Large ARCI, on the other hand, means that for some models sufficiently low χ^2_{abs} were obtained, signifying good agreement with the observations. In a sense, a threshold on ARCI is similar to a threshold on χ^2_{abs}, except that the former incorporates all aerosol mixtures simultaneously while the latter is applied mixture by mixture. Furthermore, as will be shown in section 4, ARCI is more effective than χ^2_{abs} in filtering out retrieval blunders and other obvious outliers.

Figure 1 visualizes the important steps of the method using actual MISR data. In this example, the new retrieved AOD is 0.182, whereas the V22 method gave a value of 0.174. The new retrieval uncertainty is 0.049, which is more realistic than the 0.003 uncertainty reported by the V22 algorithm. The very small uncertainty in V22 is due to the fact that only two mixtures were considered successful by passing the V22 thresholds. This example highlights a deficiency in the V22 assessment of retrieval uncertainty as the uncertainty is highly dependent on the number of passing mixtures as well as the value of the four separate thresholds used to determine which mixtures are considered successful. The new procedure eliminates the need for thresholds in determining AOD and its uncertainty, and the only threshold involved is applied to the single ARCI parameter, which is used as a retrieval quality indicator.

4. Retrieval quality assessment

In the MISR V22 retrieval algorithm several thresholds were set to filter out mixtures that do not provide a good match to the instrument observations. The threshold that provides the most strict screening in V22 is $\chi^2_{abs} \leq 2$, which is applied individually to each aerosol mixture. Because the thresholds provide an additional line of defense against clouds that were not screened by other procedures in the aerosol retrieval process, elimination of these thresholds can result in a large number of high-AOD retrievals in areas that are notorious for frequent cloud cover, but have climatologically very low AODs.

This situation is illustrated in Figure 2, which shows the average AOD for the combination of January and July of 2007 obtained with the new retrieval methodology. Vast areas of the high-latitude oceans are speckled with unrealistically high AODs, clearly indicating an issue with cloud contamination. In V22 the χ^2_{abs} and other thresholds are able to limit the occurrence of such blunders. In the new algorithm, which performs aerosol retrievals on a 4.4 km grid, in contrast to

the coarser 17.6 m grid used in V22, the problem of cloud contamination is further amplified due to closer proximity to cloud edges. Applying the same thresholds as in V22 does not fully mitigate the issue: substantially more 4.4 km retrievals remain cloud contaminated than in V22 (results not shown). Fortunately, the ARCI metric introduced in the previous section proves to be extremely effective at filtering out potentially cloud-contaminated AOD retrievals.

Figure 3a shows average AOD from 4.4 km retrievals as a function of the minimum value of χ^2_{abs}. In total about 49 million retrievals were analyzed here. After a rapid initial drop related to a similar rapid increase in sampling (Fig. 3b), the average AOD increases gradually with increasing $\min\left(\chi^2_{abs}\right)$, while the sampling continues decreasing. The AOD increase could be due to a combination of the increasing number, magnitude, and relative occurrence of cloud-contaminated, high-AOD retrievals with increasing $\min\left(\chi^2_{abs}\right)$. Based on the V22 χ^2_{abs} threshold approach, all retrievals with $\min\left(\chi^2_{abs}\right) \leq 2.0$ would have been considered successful. However, no clear justification for a threshold, either at 2.0, or any other value, is evident in the average AOD data. Choosing a value for the threshold that minimizes the average AOD would screen clouds but also potentially screen optically thick aerosol plumes, such as the heavy dust that is prevalent off the west coast of Africa. The picture looks different, however, when one considers the average AOD as a function of ARCI (Fig. 3c). After excluding the initial fluctuation for extremely small ARCI related to poor sampling, two distinct regimes in the trend of average AOD can be noticed. In the first regime, the average AOD is highly sensitive to the specific value of ARCI, characterized by a sharp decrease in AOD with increasing ARCI between about 0.03 and 0.13. This suggests that a decreasing number of cloud-contaminated, high-AOD retrievals are included in the average as the ARCI is increased. Indeed, the percentage of retrievals with AOD higher than 2.0 reaches its peak, 16%, at ARCI equal to 0.03, and decreases to about 2% when ARCI is 0.13. In the second regime, there are relatively small changes in the average AOD as ARCI increases above 0.15. The low AOD gradient in the second regime suggests a low prevalence of cloud contaminated or erroneous AODs. The retrieval count decreases slowly with increasing ARCI (Fig. 3d), indicating that the observed trends in the average AOD cannot be ascribed to a change in frequency. Conveniently, the number of screened retrievals with ARCI ≥ 0.15 is similar to the number of retrievals that do not pass the $\chi^2_{abs} \leq 2.0$ threshold. Out of about 49 million retrievals, 35.9% are below the ARCI threshold (not passing), and 37.1% are above the $\chi^2_{abs} \leq 2.0$ threshold. We set ARCI ≥ 0.15 as the value to be used as a threshold for screening retrieval blunders due to potential cloud contamination or other factors.

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 4b shows the respective retrieval count. The previous χ^2_{abs} threshold limit at 2.0 is marked with the black vertical line. All retrievals to the left of this line would have been considered successful in the V22 algorithm. This

includes a small group of high-AOD retrievals with min(χ^2_{abs}) close to 0.2 and ARCI about 0.1. Another suspicious group of retrievals with high average AODs that would have passed the previous threshold is close to min(χ^2_{abs}) = 2.0. The new ARCI threshold limit, marked with the gray horizontal line, eliminates most of the suspiciously high-AOD regions. All retrievals above the gray horizontal line are considered to be of sufficiently good quality. Of course, more complicated relationships could be investigated, but the use of ARCI as a single screening parameter proves to be highly efficient and furthermore has the advantage of simplicity. The V23 MISR aerosol product will provide the values of both ARCI and χ^2_{abs} for use in exploring custom-made cloud screenings and for other purposes.

Figure 5 presents the average AOD distribution obtained using the combination of the January and July 2007 data with retrieval screening based on the ARCI metric (ARCI≥0.15). This result is directly comparable to Fig. 2, which uses the unscreened data. The benefit of ARCI screening is readily apparent. AODs in large swaths of remote oceans are now represented by smaller and more realistic values (Witek et al., 2013). At the same time, climatologically large AODs off the coasts of Africa and South and East Asia are retained, indicating that the new screening method does not unintentionally remove all high AODs that are likely valid. The global average AOD is reduced from 0.295 for the unscreened data to 0.141 with ARCI screening. However, speckles of high AOD values are still present in many remote and cloudy regions. The majority of these retrievals are visibly cloud contaminated. This demonstrates that the ARCI screening is not ideal as some erroneous AODs pass the threshold. Increasing the threshold reduces the appearance of blunders, but also decreases the number of valid low- and moderate-AOD retrievals, reducing the overall coverage. Because the choice of setting the ARCI threshold limit at 0.15 is well supported statistically (see Figs. 3c, 4a), the remaining cloud-contaminated AOD retrievals should be addressed using another screening method. A possible approach is to employ the clear flag fraction metric discussed in Witek et al. (2013). The application of this approach to removing the remaining cloud-contaminated retrievals in the MISR V23 aerosol product will be discussed in a separate paper.

5. Statistical assessment of AOD retrieval uncertainty.

The AOD retrieval uncertainty described by Eq. 4 is a measure of the sensitivity of the algorithm to the assumed aerosol microphysical properties. This is an important factor affecting retrieval uncertainty (Li et al., 2009; Povey and Grainger, 2015), but, as mentioned in the introduction, there are many other sources of error not accounted for in this approach. Hence, the AOD uncertainty obtained from the algorithm should not be interpreted as a measure of how far the retrieved AOD deviates from the "truth". This is an important distinction that needs to be properly understood. The calculated uncertainty is purely algorithmic and depends on the initial choice of aerosol mixtures that go into the MISR SMART LUT. Any changes in the prescribed mixtures would lead to different uncertainty estimates. Furthermore, if Eq. 2 is modified such that a different goodness-of-fit metric is used, a different uncertainty result would be expected. For

these reasons, interpretation of the established uncertainty does not extend beyond the algorithm's performance. It does, however, help establish confidence intervals on the retrievals when comparing one pixel to another.

In Figure 6, AOD uncertainty is plotted as a function of AOD using the combined data from January and July of 2007. Only ARCI-screened retrievals are considered. Reported uncertainties are generally much smaller than their associated AODs. Only for very low AOD values do the uncertainties exceed the retrieved AODs. The linear fit to the data indicates that the uncertainty is about 12% of the AOD and has an offset of 0.012. This offset is much smaller than what has been traditionally discussed in the literature, specifically the 0.03 or 0.05 in the smaller and larger EE, respectively (e.g. Kahn et al., 2010). At low AODs (<0.03) the reported uncertainty is almost always below these offset levels. At higher AODs, uncertainties are often smaller than 20% or even 10% of the retrieved AODs. Note that there is always a substantial spread of AOD uncertainties at any given AOD level, often over an order of magnitude. This shows that, at least from a statistical perspective, the algorithm is capable of representing variability in retrieval confidence. For example, a retrieved AOD of 0.1 can have an uncertainty of 0.05 or of 0.005 depending on circumstances. Assigning physical meaning to a particular uncertainty value as a departure from the true value, as stated earlier, is a task that needs to be addressed separately.

A comparison of the new AOD uncertainties against their V22 predecessors reveals many similarities between the two, as evidenced in Figure 7. Recall that the V22 algorithm calculated uncertainties based on the AOD that minimized χ^2_{abs} for each mixture, while the V23 algorithm evaluates the full range of χ^2_{abs} as a function of AOD, so this agreement is not accidental. On average the V23 uncertainty is larger than that reported in V22. There does not seem to be a lower limit on uncertainties in V22, often exhibiting values below 10^{-3} whereas in V23 they mostly stay above 2×10^{-3}. The small values reported in V22 may be due to situations where only a single mixture was considered successful. Furthermore, there is discernable quantization, or clustering, of uncertainties in V22, visible as vertical striping in Fig. 7. This quantization is clearly eliminated in V23. Overall, the new AOD uncertainties appear to have a more reasonable statistical behavior compared to the uncertainties obtained in V22.

6. Conclusions and summary.

Ensemble techniques have been widely used in weather forecasting applications and climate research. They are indispensable in characterizing uncertainties and errors of highly non-linear systems, where standard error propagation techniques cannot be applied. These techniques are also useful tools for quantifying uncertainties in satellite remote retrievals of geophysical quantities (Povey and Grainger, 2015). MISR's aerosol retrieval strategy is a good example of the application of ensemble techniques to retrieval uncertainty assessment in operational data processing.

MISR's aerosol retrieval algorithm uses minimization of a cost function between observations and pre-calculated signals as a function of AOD. The spread of the cost function

around a particular AOD value is one indication of the uncertainty of the retrieved solution. Additionally, an ensemble of cost functions for different aerosol mixtures samples sensitivity of the retrieval process to the assumed aerosol optical and microphysical properties. This is one of the major sources of uncertainty in passive remote sensing of AOD. By including an ensemble of aerosol types in the retrieval approach, an algorithmic measure of AOD retrieval uncertainty that includes the impacts of measurement uncertainties, model errors, and aerosol type variability can be effectively derived using MISR data.

This study presents a new approach to determining AODs and AOD uncertainties in MISR retrievals. The new method will become operational for dark water aerosol processing in the upcoming release of V23 of the MISR aerosol product (scheduled for 2017). Unlike the V22 algorithm, the new approach eliminates several empirical thresholds. Instead, the AOD and AOD uncertainty determination relies solely on the χ^2_{abs} metric defined by Eq. 1. All considered mixtures contribute to the final result with a varying influence depending on the shape and magnitude of the associated cost functions. This approach allows for a consistent calculation of AOD and AOD uncertainty without the need for screening acceptable mixture solutions based on a complex interplay of multiple, and somewhat arbitrary, thresholds.

An unintended side effect of the new retrieval approach is an increased abundance of (mostly) cloud-contaminated, high-AOD retrievals in oceanic areas where very low aerosol concentrations are expected. Those blunders—remnants of imperfect cloud screening—were also present in V22, but many were rejected through the use of thresholds on different cost functions. They are more apparent in the V23 results due to the increase in the spatial resolution of the product from 17.6 km in V22 to 4.4 km in V23. Fortunately, an effective screening criterion has been established that filters out most cloud-contaminated retrievals. An analysis of the ARCI metric strongly suggests a specific threshold value, below which the retrievals become increasingly contaminated by clouds. Although this screening method does not eliminate all AOD outliers, it is superior to the previously used thresholds in the V22 of the MISR aerosol product. Additional cloud screening making use of the clear fraction flag with retrieval regions is built into the V23 algorithm, and will be described separately. Comparison of the new V23 algorithm results (AODs and AOD uncertainties, in particular) to other products, specifically AERONET sunphotometer measurements, will be addressed in future publications.

Acknowledgement.

The research was carried out at the Jet Propulsion Laboratory, California Institute of Technology, under a contract with the National Aeronautics and Space Administration. Support from the MISR Project is acknowledged. We thank Ralph Kahn for providing helpful comments on the manuscript.

References:

Abdou, W. A., Martonchik, J. V, Kahn, R. A., West, R. A. and Diner, D. J.: A modified linear-mixing method for calculating atmospheric path radiances of aerosol mixtures, J. Geophys. Res. Atmos., 102(D14), 16883–16888, doi:10.1029/96JD03434, 1997.

Bréon, F. M., Vermeulen, A. and Descloitres, J.: An evaluation of satellite aerosol products against sunphotometer measurements, Remote Sens. Environ., 115(12), 3102–3111, doi:10.1016/j.rse.2011.06.017, 2011.

Diner, D. J., Beckert, J. C., Reilly, T. H., Bruegge, C. J., Conel, J. E., Kahn, R. A., Martonchik, J. V., Ackerman, T. P., Davies, R., Gerstl, S. A. W., Gordon, H. R., Muller, J. P., Myneni, R. B., Sellers, P. J., Pinty, B. and Verstraete, M. M.: Multiangle Image Spectroradiometer (MISR) instrument description and experiment overview, IEEE Trans. Geosci. Remote Sens., 36(4), 1072–1087, 1998.

Diner, D. J., Martonchik, J. V., Kahn, R. A., Pinty, B., Gobron, N., Nelson, D. L. and Holben, B. N.: Using angular and spectral shape similarity constraints to improve MISR aerosol and surface retrievals over land, Remote Sens. Environ., 94(2), 155–171, doi:10.1016/j.rse.2004.09.009, 2005.

Diner, D. J., Abdou, W. A., Ackerman, T. P., Crean, K., Gordon, H. R., Kahn, R. A., Martonchik, J. V, McMuldroch, S., Paradise, S. R. and Pinty, B.: Level 2 aerosol retrieval algorithm theoretical basis, Jet Propuls. Lab. Calif. Inst. Technol., 2008.

Dubovik, O., Smirnov, a., Holben, B. N., King, M. D., Kaufman, Y. J., Eck, T. F. and Slutsker, I.: Accuracy assessments of aerosol optical properties retrieved from Aerosol Robotic Network (AERONET) Sun and sky radiance measurements, J. Geophys. Res., 105(D8), 9791, doi:10.1029/2000JD900040, 2000.

Garay, M. J., Kalashnikova, O. V. and Bull, M. A.: Development and Assessment of a High Spatial Resolution (4.4 km) MISR Aerosol Product Using AERONET-DRAGON Data, Atmos. Chem. Phys. Discuss., 1–28, doi:10.5194/acp-2016-569, 2016.

Holben, B. N., Eck, T. F., Slutsker, I., Tanré, D., Buis, J. P., Setzer, A., Vermote, E., Reagan, J. A., Kaufman, Y. J., Nakajima, T., Lavenu, F., Jankowiak, I. and Smirnov, A.: AERONET—A Federated Instrument Network and Data Archive for Aerosol Characterization, Remote Sens. Environ., 66(1), 1–16, doi:10.1016/S0034-4257(98)00031-5, 1998.

Hubanks, P. A.: MODIS Atmosphere QA Plan for Collection 006, Greenbelt, MD 20071, USA. [online] Available from: https://modis-atmos.gsfc.nasa.gov/reference_atbd.html, 2015.

Kahn, R., Banerjee, P. and McDonald, D.: Sensitivity of multiangle imaging to natural mixtures of aerosols over ocean, J. Geophys. Res., 106(D16), 18219, doi:10.1029/2000JD900497, 2001.

Kahn, R. A. and Gaitley, B. J.: An analysis of global aerosol type as retrieved by MISR, J. Geophys. Res. Atmos., 120(9), 4248–4281, doi:10.1002/2015JD023322, 2015.

Kahn, R. A., Nelson, D. L., Garay, M. J., Levy, R. C., Bull, M. A., Diner, D. J., Martonchik, J. V., Paradise, S. R., Hansen, E. G. and Remer, L. A.: MISR aerosol product attributes and statistical comparisons with MODIS, IEEE Trans. Geosci. Remote Sens., 47(12), 4095–4114, doi:10.1109/TGRS.2009.2023115, 2009.

Kahn, R. A., Gaitley, B. J., Garay, M. J., Diner, D. J., Eck, T. F., Smirnov, A. and Holben, B. N.:

Multiangle Imaging SpectroRadiometer global aerosol product assessment by comparison with the Aerosol Robotic Network, J. Geophys. Res. Atmos., 115(23), doi:10.1029/2010JD014601, 2010.

Kalashnikova, O. V and Kahn, R.: Ability of multiangle remote sensing observations to identify and distinguish mineral dust types: 2. Sensitivity over dark water, J. Geophys. Res. Atmos., 111(D11), doi:10.1029/2005JD006756, 2006.

Kalashnikova, O. V., Garay, M. J., Martonchik, J. V. and Diner, D. J.: MISR Dark Water aerosol retrievals: Operational algorithm sensitivity to particle non-sphericity, Atmos. Meas. Tech., 6(8), 2131–2154, doi:10.5194/amt-6-2131-2013, 2013.

Lee, H., Kalashnikova, O. V., Suzuki, K., Braverman, A., Garay, M. J. and Kahn, R. A.: Climatology of the aerosol optical depth by components from the Multi-angle Imaging SpectroRadiometer (MISR) and chemistry transport models, Atmos. Chem. Phys., 16(10), 6627–6640, doi:10.5194/acp-16-6627-2016, 2016.

Levy, R. C., Remer, L. A., Kleidman, R. G., Mattoo, S., Ichoku, C., Kahn, R. and Eck, T. F.: Global evaluation of the Collection 5 MODIS dark-target aerosol products over land, Atmos. Chem. Phys., 10(21), 10399–10420, doi:10.5194/acp-10-10399-2010, 2010.

Levy, R. C., Mattoo, S., Munchak, L. A., Remer, L. A., Sayer, A. M., Patadia, F. and Hsu, N. C.: The Collection 6 MODIS aerosol products over land and ocean, Atmos. Meas. Tech., 6(11), 2989–3034, doi:10.5194/amt-6-2989-2013, 2013.

Li, Z., Zhao, X., Kahn, R., Mishchenko, M., Remer, L., Lee, K. H., Wang, M., Laszlo, I., Nakajima, T. and Maring, H.: Uncertainties in satellite remote sensing of aerosols and impact on monitoring its long-term trend: A review and perspective, Ann. Geophys., 27(7), 2755–2770, doi:10.5194/angeo-27-2755-2009, 2009.

Limbacher, J. A. and Kahn, R. A.: MISR research-aerosol-algorithm refinements for dark water retrievals, Atmos. Meas. Tech., 7(11), 3989–4007, doi:10.5194/amt-7-3989-2014, 2014.

Lynch, P., Reid, J. S., Westphal, D. L., Zhang, J., Hogan, T. F., Hyer, E. J., Curtis, C. A., Hegg, D. A., Shi, Y., Campbell, J. R., Rubin, J. I., Sessions, W. R., Turk, F. J. and Walker, A. L.: An 11-year global gridded aerosol optical thickness reanalysis (v1.0) for atmospheric and climate sciences, Geosci. Model Dev., 9(4), 1489–1522, doi:10.5194/gmd-9-1489-2016, 2016.

Martonchik, J. V., Diner, D. J., Kahn, R. A., Ackerman, T. P., Verstraete, M. M., Pinty, B. and Gordon, H. R.: Techniques for the Retrieval of aerosol properties over land and ocean using multiangle data, IEEE Trans Geosci Remote Sens, 36(4), 1212–1227, 1998.

Martonchik, J. V., Diner, D. J., Crean, K. A. and Bull, M. A.: Regional aerosol retrieval results from MISR, IEEE Trans. Geosci. Remote Sens., 40(7), 1520–1531, doi:10.1109/TGRS.2002.801142, 2002.

Omar, A. H., Winker, D. M., Tackett, J. L., Giles, D. M., Kar, J., Liu, Z., Vaughan, M. A., Powell, K. A. and Trepte, C. R.: CALIOP and AERONET aerosol optical depth comparisons: One size fits none, J. Geophys. Res. Atmos., 118(10), 4748–4766, doi:10.1002/jgrd.50330, 2013.

Povey, A. C. and Grainger, R. G.: Known and unknown unknowns: Uncertainty estimation in satellite remote sensing, Atmos. Meas. Tech., 8(11), 4699–4718, doi:10.5194/amt-8-4699-2015, 2015.

Remer, L. A., Kaufman, Y. J., Tanré, D., Mattoo, S., Chu, D. A., Martins, J. V., Li, R.-R., Ichoku, C., Levy, R. C., Kleidman, R. G., Eck, T. F., Vermote, E. and Holben, B. N.: The MODIS Aerosol Algorithm, Products, and Validation, J. Atmos. Sci., 62(4), 947–973, doi:10.1175/JAS3385.1, 2005.

Remer, L. A., Kleidman, R. G., Levy, R. C., Kaufman, Y. J., Tanré, D., Mattoo, S., Martins, J. V., Ichoku, C., Koren, I., Yu, H. and Holben, B. N.: Global aerosol climatology from the MODIS satellite sensors, J. Geophys. Res. Atmos., 113(14), 1–18, doi:10.1029/2007JD009661, 2008.

Sayer, A. M., Hsu, N. C., Bettenhausen, C., Ahmad, Z., Holben, B. N., Smirnov, A., Thomas, G. E. and Zhang, J.: SeaWiFS Ocean Aerosol Retrieval (SOAR): Algorithm, validation, and comparison with other data sets, J. Geophys. Res. Atmos., 117(3), 1–17, doi:10.1029/2011JD016599, 2012.

Sayer, A. M., Hsu, N. C., Bettenhausen, C. and Jeong, M. J.: Validation and uncertainty estimates for MODIS Collection 6 "deep Blue" aerosol data, J. Geophys. Res. Atmos., 118(14), 7864–7872, doi:10.1002/jgrd.50600, 2013.

Sayer, A. M., Hsu, N. C., Bettenhausen, C., Lee, J., Redemann, J., Schmid, B. and Shinozuka, Y.: Extending "deep Blue" aerosol retrieval coverage to cases of absorbing aerosols above clouds: Sensitivity analysis and first case studies, J. Geophys. Res. Atmos., 121(9), 4830–4854, doi:10.1002/2015JD024729, 2016.

Shi, Y., Zhang, J., Reid, J. S., Holben, B., Hyer, E. J. and Curtis, C.: An analysis of the collection 5 MODIS over-ocean aerosol optical depth product for its implication in aerosol assimilation, Atmos. Chem. Phys., 11(2), 557–565, doi:10.5194/acp-11-557-2011, 2011.

Shi, Y., Zhang, J., Reid, J. S., Hyer, E. J. and Hsu, N. C.: Critical evaluation of the MODIS Deep Blue aerosol optical depth product for data assimilation over North Africa, Atmos. Meas. Tech., 6(4), 949–969, doi:10.5194/amt-6-949-2013, 2013.

Shi, Y., Zhang, J., Reid, J. S., Liu, B. and Hyer, E. J.: Critical evaluation of cloud contamination in the MISR aerosol products using MODIS cloud mask products, Atmos. Meas. Tech., 7(6), 1791–1801, doi:10.5194/amt-7-1791-2014, 2014.

Witek, M. L., Garay, M. J., Diner, D. J. and Smirnov, A.: Aerosol optical depths over oceans: A view from MISR retrievals and collocated MAN and AERONET in situ observations, J. Geophys. Res. Atmos., 118(22), 12620–12633, doi:10.1002/2013JD020393, 2013.

Zhang, J. and Reid, J. S.: MODIS aerosol product analysis for data assimilation: Assessment of over-ocean level 2 aerosol optical thickness retrievals, J. Geophys. Res. Atmos., 111(22), 1–17, doi:10.1029/2005JD006898, 2006.

Zhang, J. and Reid, J. S.: An analysis of clear sky and contextual biases using an operational over ocean MODIS aerosol product, Geophys. Res. Lett., 36(15), 1–6, doi:10.1029/2009GL038723, 2009.

Zhang, J. and Reid, J. S.: A decadal regional and global trend analysis of the aerosol optical depth using a data-assimilation grade over-water MODIS and Level 2 MISR aerosol products, Atmos. Chem. Phys., 10(22), 10949–10963, doi:10.5194/acp-10-10949-2010, 2010.

Zhang, J., Reid, J. S., Westphal, D. L., Baker, N. L. and Hyer, E. J.: A system for operational aerosol optical depth data assimilation over global oceans, J. Geophys. Res. Atmos., 113(10), 1–13, doi:10.1029/2007JD009065, 2008.

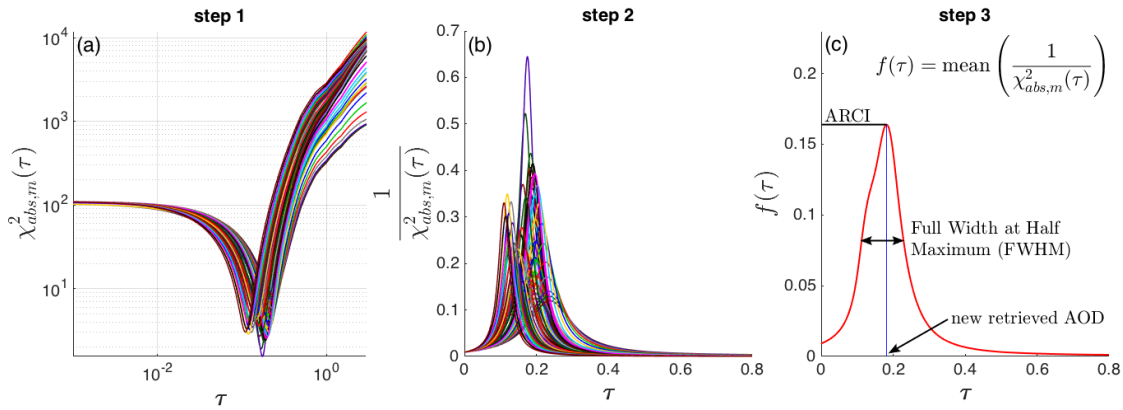


Figure 4 Example of calculation steps performed in the new methodology for determining AOD and its uncertainty. (a) χ^2_{abs} values for 74 MISR mixtures as a function of AOD (τ) (Eq. 1); (b) inverse (reciprocal) values for the 74 mixtures; and (c) inverse residuals averaged over all mixtures (Eq. 2), with the new retrieved AOD, ARCI, and FWHM indicated on the distribution. The x-axis scale is logarithmic in panel (a) for a better visualization of the cost function at low τ.

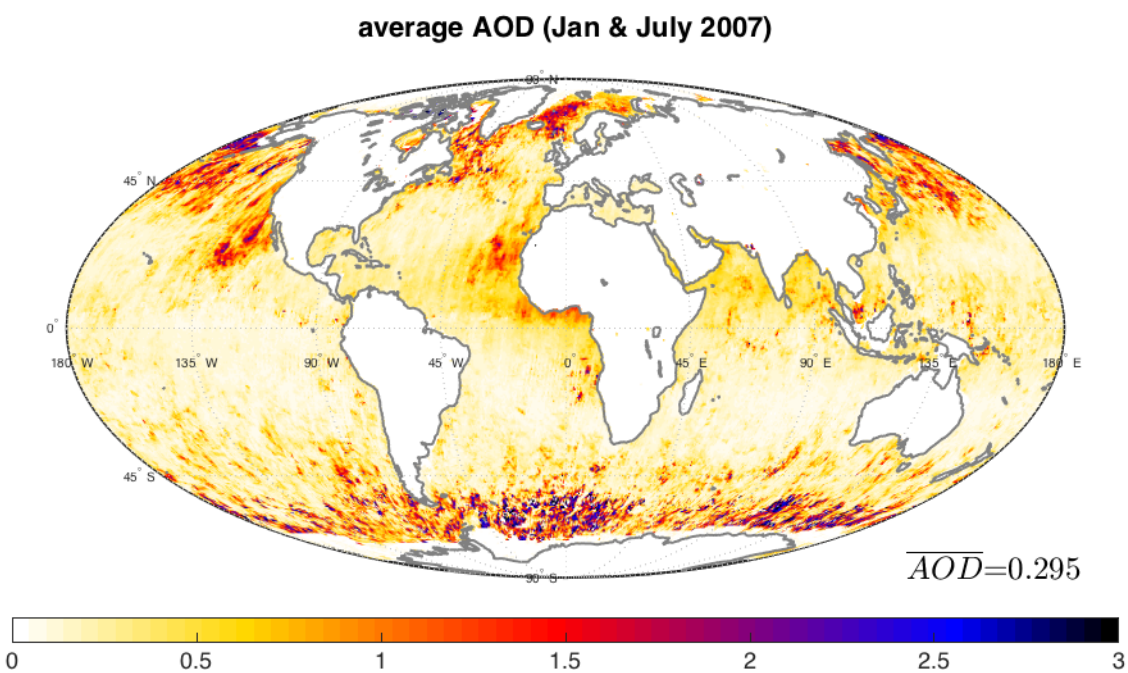


Figure 5 Average AOD obtained using unscreened data from January and July of 2007. The data are mapped to a 0.5º × 0.5º grid. High AOD values over remote oceans indicate issues with cloud contamination in the retrieval process.

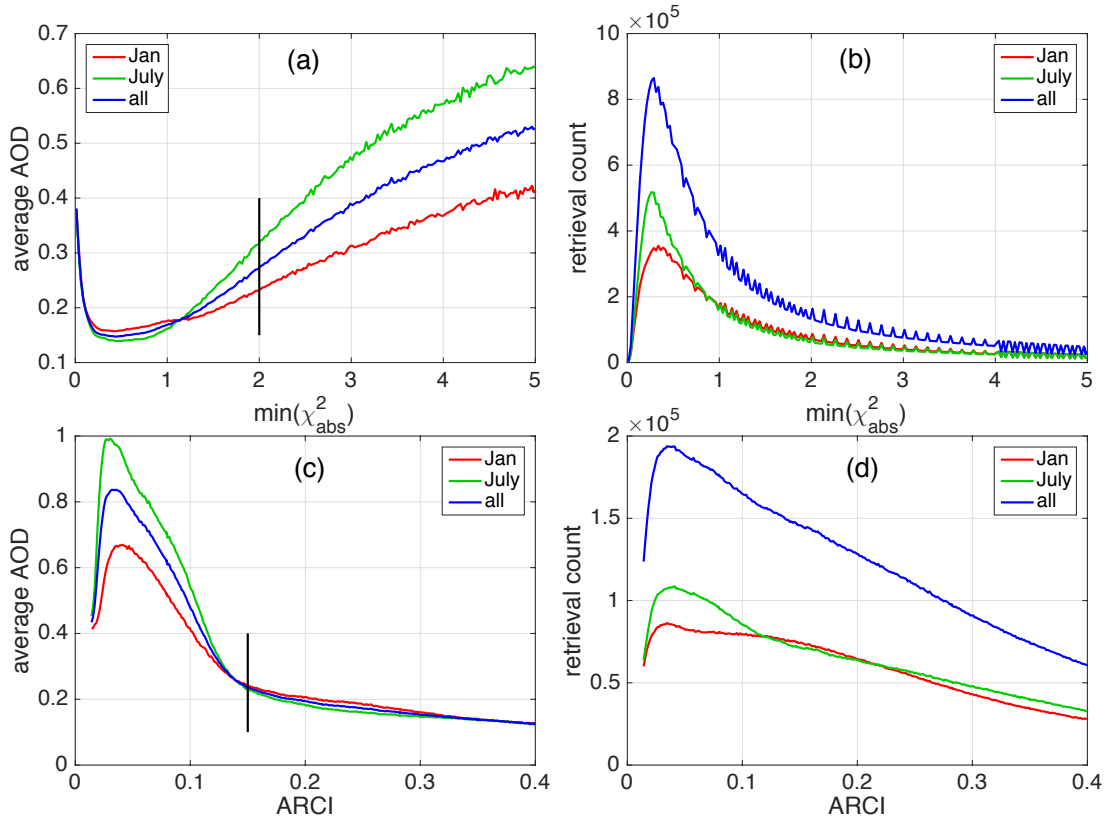


Figure 3 Histograms of average AOD as a function of (a) min(χ^2_{abs}), and (c) ARCI for January 2007, July 2007, and the two months combined. Panels (b) and (d) are histograms of retrieval counts corresponding to min(χ^2_{abs}) and ARCI values, respectively.

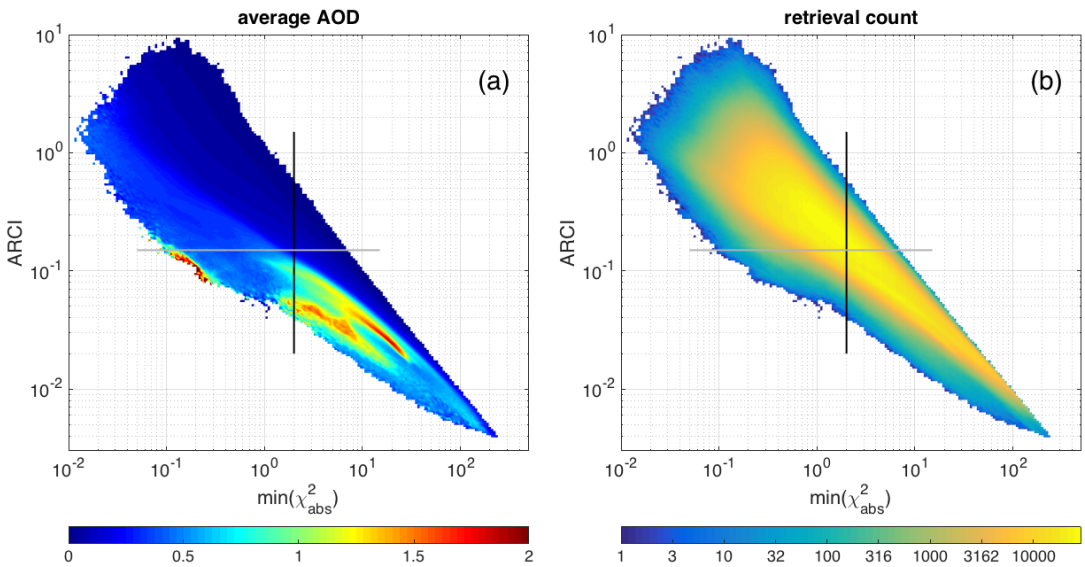


Figure 4 (a) average AOD as a function of ARCI and min(χ^2_{abs}) for the combined months of January and July of 2007, (b) retrieval count for the data plotted in panel (a).

5

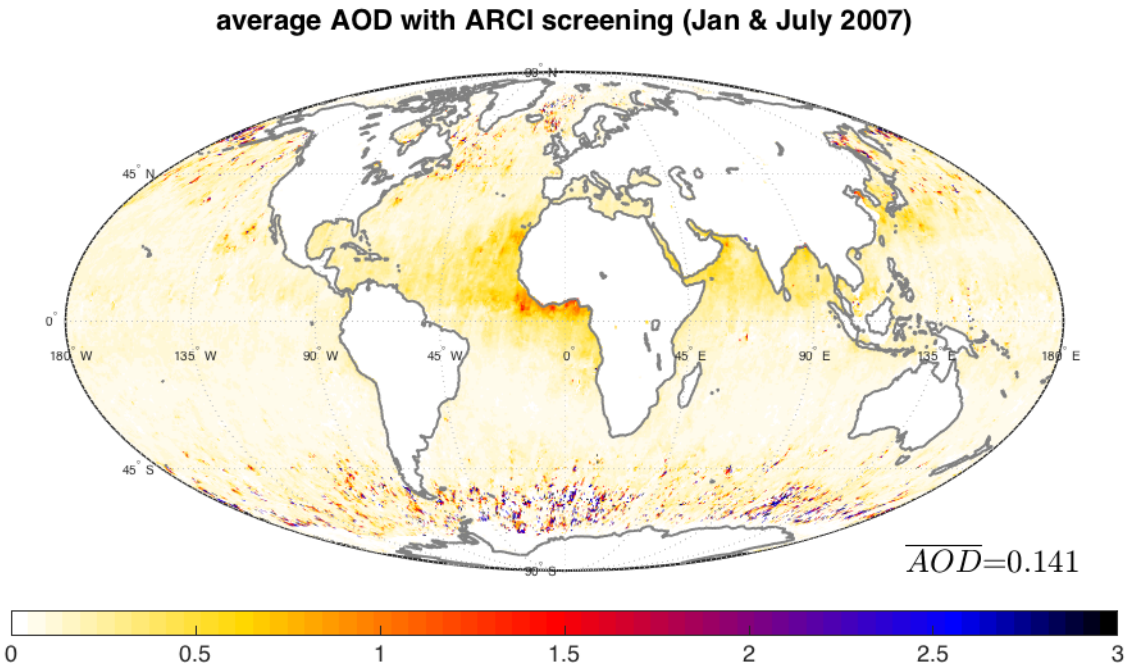


Figure 5 Average AOD distribution with ARCI≥0.15 screening for the combination of January and July 2007. The data are mapped to a 0.5º × 0.5º grid. There is substantial improvement in the global distribution of mean AODs when compared to the unscreened data in Fig. 2. However, some residual high AOD values remain over remote oceans. They can be further screened using other approaches.

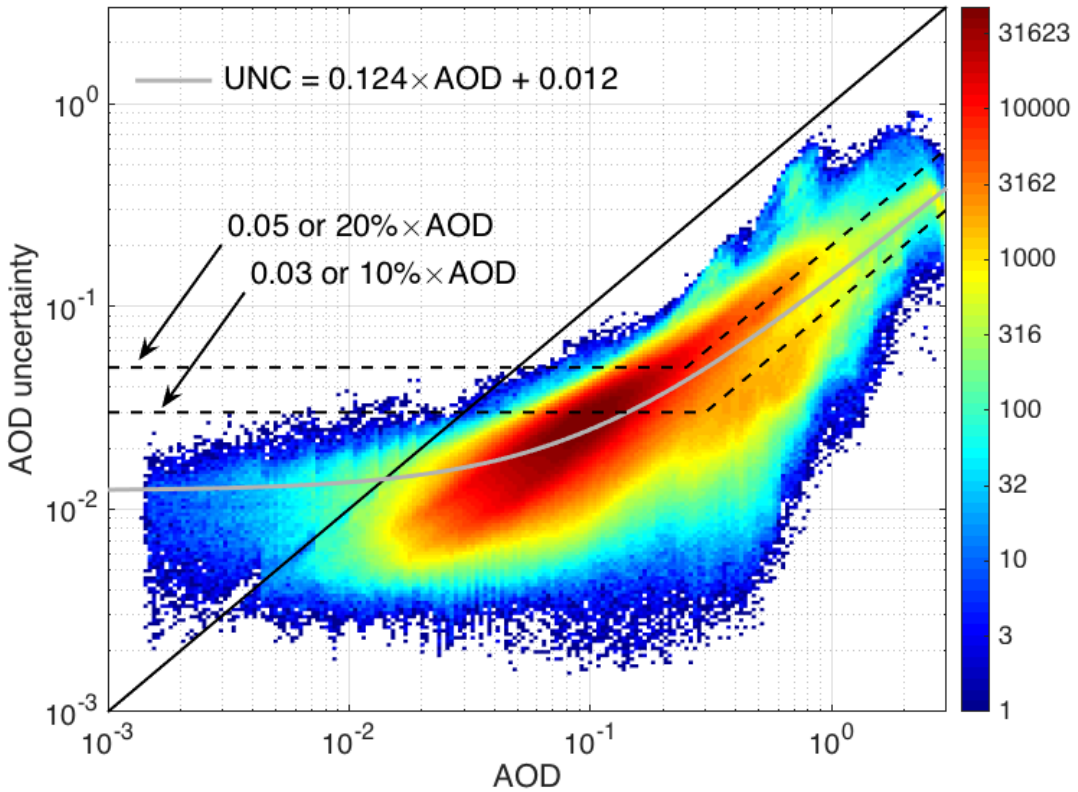


Figure 6 Density plot of AOD uncertainty in log-log space as a function of AOD for the combined January and July 2007 data with ARCI screening. The black line is the 1-to-1 line, included as a visual guide to illustrate that, over most of the AOD range, the uncertainties are smaller than the AOD values themselves. The gray line is a linear fit to the data. Two dashed lines represent two arbitrary uncertainty envelopes.

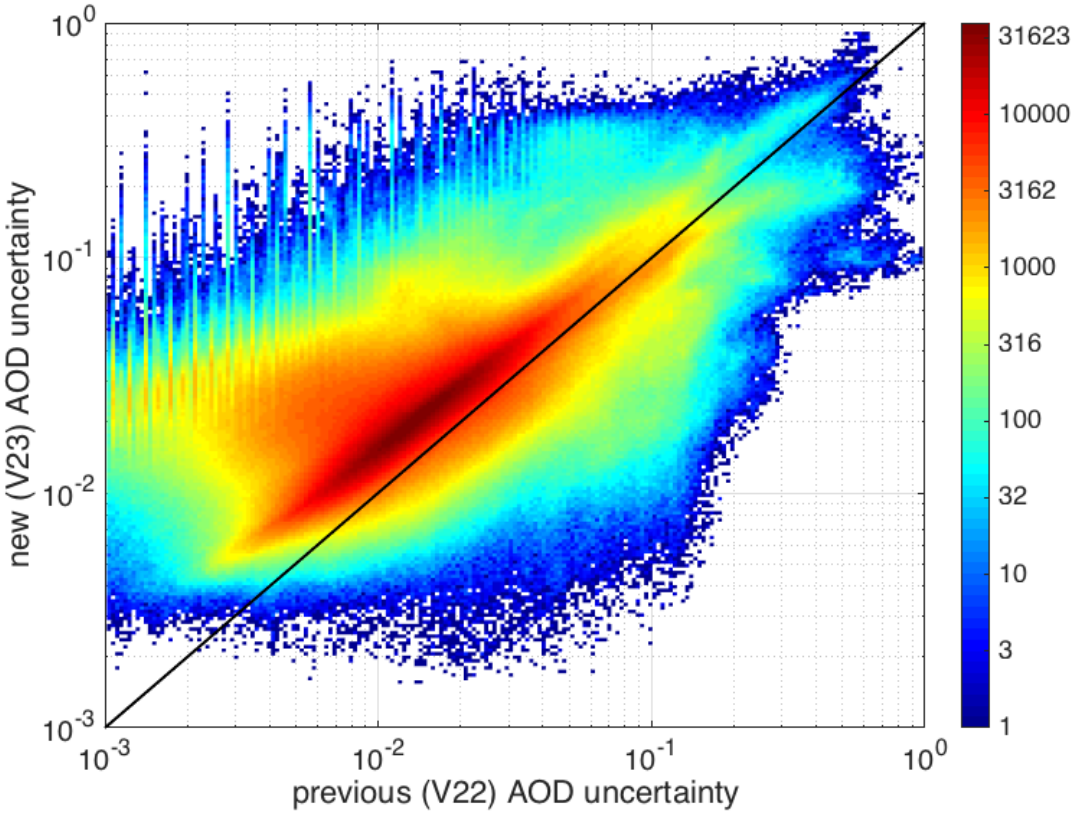


Figure 7 Density plot showing comparison between the previous (V22) AOD uncertainty and the new (V23) AOD uncertainty. The black line is the 1-to-1 line.